

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 09:32:25 ; Search time 57 Seconds
(without alignments)

3316.215 Million cell updates/sec

Title: US-10-086-913-2

Perfect score: 3517

Sequence: 1 MGSSHHHHHSSGLVPRGSH.....EAHMGSSSSSSTPTGSGC 669

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq_29Jan04:.*
1: Geneseqp1980s:.*
2: Geneseqp1990s:.*
3: Geneseqp2000s:.*
4: Geneseqp2001s:.*
5: Geneseqp2002s:.*
6: Geneseqp2003as:.*
7: Geneseqp2003bs:.*
8: Geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3517	100.0	669	5 ABB08420	Abb08420 Catalytic
2	3517	100.0	669	7 ADD06430	Add06430 Variant T
3	3360	95.5	1060	2 AAY01540	Aay01540 Trypanoso
4	3269	92.9	642	2 AAY01541	Aay01541 Alpha(2-3
5	3110.5	88.4	666	7 ADC21501	Adc21501 T. cruzi
6	2681	76.2	1162	7 ADC21533	Adc21533 T. cruzi
7	1067	30.3	200	7 ADC21529	Adc21529 T. cruzi
8	1037	29.5	200	7 ADC21530	Adc21530 T. cruzi
9	903	25.7	1003	3 AAY44454	Aay44454 T. cruzi
10	903	25.7	1006	3 AAY44455	Aay44455 Modified
11	884	25.1	166	2 AAR42017	Aar42017 Trans-sia
12	884	25.1	166	2 AAR42014	Aar42014 Protein w
13	854	24.3	166	2 AAR42015	Aar42015 Trans-sia
14	825	23.5	166	2 AAR42016	Aar42016 TCNA Tran
15	631.5	18.0	618	2 AAW26541	Aaw26541 Trypanoso
16	630.5	17.9	618	2 AAY23320	Aay23320 Trpanosom
17	388	11.0	382	7 ABB80240	Abb80240 Wildtype
18	384	10.9	382	7 ABB80239	Abb80239 Synthetic
19	319.5	9.1	433	2 AAW48868	Aaw48868 Pasteurel
20	241	6.9	544	2 AAW02207	Aaw02207 Bacteroid
21	234.5	6.7	396	2 AAR47061	Aar47061 Bacteroid
22	201	5.7	800	6 ABU02167	Abu02167 S. pneumo
23	201	5.7	800	6 ABP81446	Abp81446 Streptoco
24	184	5.2	492	2 AAY34503	Aay34503 Porphyorm
25	184	5.2	540	2 AAY34376	Aay34376 Porphyorm

ALIGNMENTS

RESULT 1

ABB08420

ID ABB08420 standard; protein; 669 AA.

AC ABB08420;

XX 01-JUL-2002 (first entry)

XX Catalytic trans-sialidase unit of T. cruzi amino acid sequence.

XX Mycoplasma associated disease; cell proliferation; trans-sialidase;
enzyme; atherosclerotic vascular disease; malignancy; sialic acid;
antithrombotic; antibacterial; antiviral; anti-HIV; cytostatic;
KW vasotropic; ovarian carcinoma; breast cancer; prostate cancer;
colon cancer; lung cancer; leukaemia; HIV; human immunodeficiency virus;
chlamydia; PCR primer.

XX Trypanosoma cruzi.

OS Synthetic.

XX WO200202050-A2.

XX WO200202050-A2.

XX 10-JAN-2002.

XX 03-JUL-2001; 2001WO-BR000083.

XX 03-JUL-2000; 2000BR-00002989.

XX (HIGU/) HIGUCHI M D L.

XX (SCHE/) SCHENKMAN S.

XX Higuchi MDL, Schenkman S;

XX WPI; 2002-154675/20.

XX N-PSDB; ABA98876.

XX Composition useful for treatment of mycoplasma infection and diseases

XX associated with cell proliferation e.g. malignancy or with co-infection

XX with another microbe, comprises agent inhibiting sialic acid-mediated

XX attachment of mycoplasma.

XX Claim 6; Fig 26; 63pp; English.

XX The invention relates to a composition useful for treating or preventing

XX mycoplasma infection in a subject suffering from a disorder characterised

XX by increased cell proliferation or by co-infection with a second microbe,

XX comprising an agent that prevents or inhibits sialic acid-mediated

XX attachment of mycoplasma to the subject's cells. The activity of

compositions of the invention may be described as; antiatherosclerotic, antibacterial, antiviral, anti-HIV, cytostatic and vasotropic. The compositions are useful to treat diseases associated with undesirable cell proliferation, such as atherosclerotic vascular disease and malignancy, by reducing or preventing mycoplasma infection. Examples of malignancies include; ovarian carcinoma, breast cancer, prostate cancer, colon cancer, lung cancer and leukaemia. They are also useful to treat diseases associated with infection with other infectious organisms co-occurring with mycoplasma (and typically increasing the virulence of both pathogens), especially human immunodeficiency virus or chlamydia species. They can be used to treat such diseases in humans or other animals, and can be administered in conjunction with conventional agents e.g. anti-platelet or chemotherapeutic agents. The current sequence represents the catalytic trans-sialidase unit of T. cruzi amino acid sequence

XX Sequence 669 AA;

Query Match 100.0%; Score 3517; DB 5; Length 669;
Best Local Similarity 100.0%; Pred. No. 4.9e-290;
Matches 669; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSSHHHHHSSGLVPRGSHMAPGSSRVLEFKROSSKVPPEKGGKVTERRVHVSFRPALV 60
DB 1 MGSSHHHHHSSGLVPRGSHMAPGSSRVLEFKROSSKVPPEKGGKVTERRVHVSFRPALV 60

QY 61 NVDGWVAIAADARYETSDNLSLIDTVAKYSYDDGETWETQIAIKNSRASSVSRVVDPTVI 120
DB 61 NVDGWVAIAADARYETSDNLSLIDTVAKYSYDDGETWETQIAIKNSRASSVSRVVDPTVI 120

QY 121 VGNKLYLVGYSNRSRYTSHGDARDWDILLAVGEVTKSTAGGKITASKWGPSVSLK 180
DB 121 VGNKLYLVGYSNRSRYTSHGDARDWDILLAVGEVTKSTAGGKITASKWGPSVSLK 180

QY 181 EFPFAEMEGMTNPLGAGVAIVASNGNLVYPQVTNKKKQVPSKIFYSDESGKTKWFG 240
DB 181 EFPFAEMEGMTNPLGAGVAIVASNGNLVYPQVTNKKKQVPSKIFYSDESGKTKWFG 240

QY 241 EGRDFFGSGSEPALEWEGKLIINTRVYRRRLVYESSDMGNSWVEAVGTLRSVMGSPSKS 300
DB 241 EGRDFFGSGSEPALEWEGKLIINTRVYRRRLVYESSDMGNSWVEAVGTLRSVMGSPSKS 300

QY 301 NOPGSQSFATVTEGHRVMLFTHPLNPKGRWLRDLNLTQNRINYVGVISIGDENS 360
DB 301 NOPGSQSFATVTEGHRVMLFTHPLNPKGRWLRDLNLTQNRINYVGVISIGDENS 360

QY 361 AYSSVLYKDKLYCLHEINSNEVYSLVFARLVGELRIIKSVLQSWKNWDSHLSICTPAD 420
DB 361 AYSSVLYKDKLYCLHEINSNEVYSLVFARLVGELRIIKSVLQSWKNWDSHLSICTPAD 420

QY 421 PAASSSERGGCPAVTTVGLVFLSHSATKTEWEDAYRCVNASTANAEVPGNGLKFAVG 480
DB 421 PAASSSERGGCPAVTTVGLVFLSHSATKTEWEDAYRCVNASTANAEVPGNGLKFAVG 480

QY 481 GALWPVSQOQONRYHPANHAFTLVASVTTHVPSVASPLLGASLDSSGKLLGLSYDE 540
DB 481 GALWPVSQOQONRYHPANHAFTLVASVTTHVPSVASPLLGASLDSSGKLLGLSYDE 540

QY 541 KHWQPIYGTPTVPTGSEWEMGRYHVLTWANKIGSVYIDGEPFGSGQTVVPDGRTPD 600
DB 541 KHWQPIYGTPTVPTGSEWEMGRYHVLTWANKIGSVYIDGEPFGSGQTVVPDGRTPD 600

QY 601 ISHFYVGGYGRSDMPTISHVTVNNVLLYNRLNAEIRTLFLSQDLIGTEAHMGSSSGSS 660
DB 601 ISHFYVGGYGRSDMPTISHVTVNNVLLYNRLNAEIRTLFLSQDLIGTEAHMGSSSGSS 660

QY 661 ERSTPGSGC 669
DB 661 ERSTPGSGC 669

RESULT 2
ADD06430
ID ADD06430 standard; protein; 669 AA.

XX AC ADD06430;
XX DT 01-JAN-2004 (first entry)
XX DB Variant T. cruzi trans-sialidase enzyme.
XX KW Mycoplasma infection; cell proliferation; co-infection; parasite;
KW respiratory epithelium; urogenital tract; infection; AIDS; autoimmune;
KW sialic acid-mediated attachment; antibiotic; neuraminidase;
KW trans-sialidase; neoplasia; neoplastic cell; apoptosis;
KW atherosclerotic vascular disease; malignant disease; tumour;
KW human immunodeficiency virus; HIV; Chlamydia; antibacterial;
KW antiarteriosclerotic; cytostatic; anti-HIV; enzyme.
XX OS Synthetic.
OS Trypanosoma cruzi.
XX PN US2003124109-A1.
XX PD 03-JUL-2003.
XX PF 01-MAR-2002; 2002US-00086913.
XX PR 03-JUL-2000; 2000BR-00002989.
XX PR 03-JUL-2001; 2001BR-00002648.
XX PR 03-JUL-2001; 2001WO-BR000083.
XX PA (HIGU/) HIGUCHI M D L.
PA (SCHE/) SCHENKMAN S.
XX PI Higuchi MDL, Schenkman S;
XX WPI; 2003-810968/76.
XX N-PSDB; ADD06429.
PT Use of an agent that prevents or inhibits Mycoplasma infection, for
PT manufacturing a medicament for treating or preventing a disorder
PT associated with increased cell proliferation, e.g. atherosclerotic
PT vascular disease or malignancy.
XX Claim 6; SEQ ID NO 2; 32pp; English.
XX The invention discloses the use of an agent that prevents or inhibits
CC Mycoplasma infection for manufacturing a medicament for treating a
CC disorder associated with increased cell proliferation or a co-infection
CC with mycoplasma and a second microbe. Mycoplasmas are parasites of the
CC respiratory epithelium and urogenital tract. Infections are typically
CC asymptomatic but they seem to be co-factors in diseases such as AIDS and
CC in sequelae after mycoplasma infections having an autoimmune basis. The
CC agent prevents or inhibits sialic acid-mediated attachment of mycoplasma
CC to cells of the subject and is an antibiotic or an enzyme having an
CC activity consisting of neuraminidase and/or trans-sialidase activity. The
CC enzyme is derived from a trypanosoma cruzi microorganism, where the
CC enzyme is a native or a recombinant enzyme. Results showed that trans-
CC sialidase is effective as a drug in the treatment of neoplasia, removing
CC mycoplasmas from the neoplastic cells leading to their apoptosis. The
CC composition or the agent that prevents or inhibits mycoplasma infection
CC is useful for manufacturing a medicament for treating or preventing a
CC disorder associated with increased cell proliferation, e.g.
CC atherosclerotic vascular disease or malignant disease (tumour), or a
CC disease associated with co-infection with mycoplasma and a second microbe
CC such as human immunodeficiency virus or a Chlamydia microbe. The sequence
CC presented is the variant T. cruzi trans-sialidase enzyme.
XX Sequence 669 AA;

Query Match 100.0%; Score 3517; DB 7; Length 669;
Best Local Similarity 100.0%; Pred. No. 4.9e-290;
Matches 669; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSSHHHHHSSGLVPRGSHMAPGSSRVLEFKROSSKVPPEKGGKVTERRVHVSFRPALV 60
|||||

Db 1 MGSSHHHHHSGLVPRGSHMARGSRVLFKQSSKVPFEGKGVTERVHVSFRLPALV 60
QY 61 NVDGVVVAIADARYETSDNLSLIDTVAKYSDVDDGETWETQIAIKNSRASSVSRRVDP TVI 120
Db 61 NVDGVVVAIADARYETSDNLSLIDTVAKYSDVDDGETWETQIAIKNSRASSVSRRVDP TVI 120
QY 121 VKGNKLYLVGVSNSRSRYTSHGDARDWDILLAVGEVTKTAGGKITASIKWGSVPVSLK 180
Db 121 VKGNKLYLVGVSNSRSRYTSHGDARDWDILLAVGEVTKTAGGKITASIKWGSVPVSLK 180
QY 181 EFPFAEMEGHMTNQFLGGAGVAIVASNGNLVYPVQVTKKKQVFSKIFYSEDEGKTKWFG 240
Db 181 EFPFAEMEGHMTNQFLGGAGVAIVASNGNLVYPVQVTKKKQVFSKIFYSEDEGKTKWFG 240
QY 241 EGRSDFGCSFPVALEWEGKLIINTRDYRRRLVYESSDMGNSVAVGTLRSRVWGPSKPS 300
Db 241 EGRSDFGCSFPVALEWEGKLIINTRDYRRRLVYESSDMGNSVAVGTLRSRVWGPSKPS 300
QY 301 NQPGSSSFTAVTIEGRVMLFTHPLNFKGRWLRDRNLNLTQNRINYNGVQVSIIGDENS 360
Db 301 NQPGSSSFTAVTIEGRVMLFTHPLNFKGRWLRDRNLNLTQNRINYNGVQVSIIGDENS 360
QY 361 AYSSVLYKDKLYCLHEINSNEVYSLVPAELVGELEIRIISVLOSQWKNWDSHLSSICTPAD 420
Db 361 AYSSVLYKDKLYCLHEINSNEVYSLVPAELVGELEIRIISVLOSQWKNWDSHLSSICTPAD 420
QY 421 PAASSSERGCGPAAVTVTVGLVFLSHSATKTEWEDAYRCVNASTANAERVPNGLKFAVG 480
Db 421 PAASSSERGCGPAAVTVTVGLVFLSHSATKTEWEDAYRCVNASTANAERVPNGLKFAVG 480
QY 481 GALWPSVQSQONQRYHFANHAFTLVASVTHIEVPSVASPLLGASLSSGGKLLGLSYDE 540
Db 481 GALWPSVQSQONQRYHFANHAFTLVASVTHIEVPSVASPLLGASLSSGGKLLGLSYDE 540
QY 541 KHQWQPIYGSTPTPTGSEWEMGRYHVLTWANKIGSVYIDGPELGSGQTVVDPDGRTPD 600
Db 541 KHQWQPIYGSTPTPTGSEWEMGRYHVLTWANKIGSVYIDGPELGSGQTVVDPDGRTPD 600
QY 601 ISHFYVGGYGRSDMPTISHVTNNVLLYNRQLNAEIRTLFLSQDLIGTEAHMGSSSGSS 660
Db 601 ISHFYVGGYGRSDMPTISHVTNNVLLYNRQLNAEIRTLFLSQDLIGTEAHMGSSSGSS 660
QY 661 ERSTPGSGC 669
Db 661 ERSTPGSGC 669

RESULT 3

AA01540
ID AA01540 standard; protein; 1060 AA.

AC AA01540;

XX 15-JUN-1999 (first entry)

DE Trypanosoma cruzi alpha(2-3) trans-sialidase amino acid sequence.

DE Alpha(2-3)trans-sialidase; sialyl-oligosaccharide; sialyllactose;
KW cheese processing waste strain; (2-3)sialyllactose; gastric ulcer;
KW duodenal ulcer; arthritis; enterotoxin.

OS Trypanosoma cruzi.

XX WO9908511-A1.

XX 25-FEB-1999.

XX 13-AUG-1998; 98WO-US016756.

XX 14-AUG-1997; 97US-00911393.

XX (NEOS-) NEOSE TECHNOLOGIES INC.

XX

PI Pelletier M, Barker WA, Hakes DJ, Zopf DA;
XX WPI; 1999-190079/16.
DR N-PSDB; AAX26611.
XX
PT Production of sialyl-oligosaccharides, particularly sialyl-lactose - by
PT treating a dairy source such as a cheese processing waste stream with an
PT alpha (2-3) trans-sialidase.
XX
PS Disclosure; Fig 2; 84pp; English.
XX
XX The present sequence represents Trypanosoma cruzi alpha(2-3)trans-
XX sialidase. The protein is used in the method of the invention to produce
XX sialyl-oligosaccharides, particularly sialyllactose, which are produced
XX by treating a dairy source such as a cheese processing waste strain with
XX an alpha (2-3) trans-sialidase. The method can be used for producing
XX sialyl-oligosaccharides, such as (2-3)sialyllactose for pharmaceutical
XX use. (2-3)sialyllactose has been shown to neutralise enterotoxins of
XX various pathogenic microbes including E. coli, Vibrio cholerae and
XX Salmonella. It has also been shown that alpha(2-3)(2-3)sialyllactose
XX (alpha-Neu5Ac-(2-3)-Gal-beta-(1-4)-Glc) interferes with colonisation of
XX Helicobacter pylori and thereby prevents or inhibits gastric and duodenal
XX ulcers. (2-3)sialyllactose has additionally been proposed to inhibit
XX immune complex formation by disrupting occupancy of the Fc carbohydrate
XX binding site on IgG and to be useful in treating arthritis
XX
XX Sequence 1060 AA;
Query Match 95.5%; Score 3360; DB 2; Length 1060;
Best Local Similarity 99.2%; Pred. No. 2.3e-276;
Matches 642; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 21 MAPGSRVLEFRKQSSKVPFEGKGVTERVHVSFRLPALVNDGVVVAIADARYETSDN 80
Db 33 LAPGSRVLEFRKQSSKVPFEGKGVTERVHVSFRLPALVNDGVVVAIADARYETSDN 92
QY 81 SLIDTVAKYSDVDDGETWETQIAIKNSRASSVSRRVDP TVIKGNKLYLVGVSNSRSY 140
Db 93 SLIDTVAKYSDVDDGETWETQIAIKNSRASSVSRRVDP TVIKGNKLYLVGVSNSRSY 152
QY 141 TSHGDARDWDILLAVGEVTKTAGGKITASIKWGSVPVSLKEFPFAEMEGHMTNQFLGGAG 200
Db 153 TSHGDARDWDILLAVGEVTKTAGGKITASIKWGSVPVSLKEFPFAEMEGHMTNQFLGGAG 212
QY 201 VAIASNGNLVYPVQVTKKKQVFSKIFYSEDEGKTKWFGESRDFGCSFPVALEWEGKL 260
Db 213 VAIASNGNLVYPVQVTKKKQVFSKIFYSEDEGKTKWFGESRDFGCSFPVALEWEGKL 272
QY 261 IINTRDYRRRLVYESSDMGNSVAVGTLRSRVWGPSKNSQPGSSSFTAVTIEGRV 320
Db 273 IINTRDYRRRLVYESSDMGNSVAVGTLRSRVWGPSKNSQPGSSSFTAVTIEGRV 332
QY 321 LFTHPLNFKGRWLRDRNLNLTQNRINYNGVQVSIIGDENSAYSSVLYKDKLYCLHEINS 380
Db 333 LFTHPLNFKGRWLRDRNLNLTQNRINYNGVQVSIIGDENSAYSSVLYKDKLYCLHEINS 392
QY 381 NEVYSLVPAELVGELEIRIISVLOSQWKNWDSHLSSICTPADPAASSERGCGPAAVTVGLV 440
Db 393 NEVYSLVPAELVGELEIRIISVLOSQWKNWDSHLSSICTPADPAASSERGCGPAAVTVGLV 452
QY 441 GFLSHSATKTEWEDAYRCVNASTANAERVPNGLKFAVGCGGALWPVVSQQQONQRYHFANH 500
Db 453 GFLSHSATKTEWEDAYRCVNASTANAERVPNGLKFAVGCGGALWPVVSQQQONQRYHFANH 512
QY 501 APTLVASVTHIEVPSVASPLLGASLSSGGKLLGLSYDEKHKWQPIYGSTPTPTPGSWE 560
Db 513 APTLVASVTHIEVPSVASPLLGASLSSGGKLLGLSYDEKHKWQPIYGSTPTPTPGSWE 572
QY 561 MGKRYHVLTWANKIGSVYIDGPELGSGQTVVDPDGRTPDISHFYVGGYGRSDMPTISHV 620
Db 573 MGKRYHVLTWANKIGSVYIDGPELGSGQTVVDPDGRTPDISHFYVGGYGRSDMPTISHV 632
QY 621 TVNNVLLYNRQLNAEIRTLFLSQDLIGTEAHMGSSSGSSSERSTPGS 667

```

Db      633 TVNNVLLYNRLNAEIRTLFLSQDLIGTEAHWGSSGSSAHTSTPST 679
|||||
RESULT 4
AAV01541
ID      AAY01541 standard; protein; 642 AA.
AC      AAY01541;
XX
DT      15-JUN-1999 (first entry)
XX
DE      Alpha(2-3) trans-sialidase sequence lacking amino acid repeats.
XX
KW      Alpha(2-3)trans-sialidase; sialyl-oligosaccharide; sialyllactose;
KW      cheese processing waste strain; (2-3)sialyllactose; gastric ulcer;
KW      duodenal ulcer; arthritis; enterotoxin.
XX
OS      Trypanosoma cruzi.
XX
PN      WO9908511-A1.
XX
PD      25-FEB-1999.
XX
PF      13-AUG-1998; 98WO-US016756.
XX
PR      14-AUG-1997; 97US-00911393.
XX
PA      (NEOS-) NEOSE TECHNOLOGIES INC.
XX
PI      Pelletier M, Barker WA, Hakes DJ, Zopf DA;
XX
DR      WPI; 1999-190079/16.
DR      N-PSDB; AAX26612.
XX
PT      Production of sialyl-oligosaccharides, particularly sialyl-lactose - by
PT      treating a dairy source such as a cheese processing waste stream with an
PT      alpha (2-3) trans-sialidase.
XX
PS      Disclosure; Fig 4; 84pp; English.
XX
CC      The present sequence represents a functional Trypanosoma cruzi alpha(2-
CC      3)trans-sialidase which lacks amino acid repeats. The protein is used in
CC      the method of the invention to produce sialyl-oligosaccharides,
CC      particularly sialyllactose, which are produced by treating a dairy
CC      such as a cheese processing waste strain with an alpha (2-3) trans-
CC      sialidase. The method can be used for producing sialyl-oligosaccharides,
CC      such as (2-3)sialyllactose for pharmaceutical use. (2-3)sialyllactose
CC      has been shown to neutralise enterotoxins of various pathogenic microbes
CC      including E. coli, Vibrio cholerae and Salmonella. It has also been shown
CC      that alpha(2-3)(2-3)sialyllactose (alpha-Neu5Ac-(2-3)-Gal-beta-(1-4)-Glc)
CC      interferes with colonisation of Helicobacter pylori and thereby prevents
CC      or inhibits gastric and duodenal ulcers. (2-3)sialyllactose has
CC      additionally been proposed to inhibit immune complex formation by
CC      disrupting occupancy of the Fc carbohydrate binding site on IgG and to be
CC      useful in treating arthritis
XX
SQ      Sequence 642 AA;

Query Match
Best Local Similarity 92.9%; Score 3269; DB 2; Length 642;
Matches 626; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY      21 MAPGSSRVLEFKQSSKVPFEKGGKTVRVHSHFRPALVNVVDGYMVAIADARYETSDN 80
Db      :|||||
QY      2 LAPGSSRVLEFKQSSKVPFEKGGKTVRVHSHFRPALVNVVDGYMVAIADARYETSDN 61
QY      81 SLIDTVAKYSVDGETWETQIAIKNSRASSVSRVVDPTVIKGNKLYLVGSGNSRSYW 140
Db      :|||||
QY      62 SLIDTVAKYSVDGETWETQIAIKNSRASSVSRVVDPTVIKGNKLYLVGSGNSRSYW 121
QY      141 TSHGDARDWDILLAVGEVTKSTAGKITASIKWGPSVSLKEFFPAEMEGMHTNQFLGGAG 200
|||||

```

```

Db      122 TSHGDARDWDILLAVGEVTKSTAGKITASIKWGPSVSLKEFFPAEMEGMHTNQFLGGAG 181
QY      201 VAIVASNGNLVYPVQVTNKKQVFSKIFYSDESGKTKWKGEGSDRGCCSEPVALEWEGKL 260
Db      182 VAIVASNGNLVYPVQVTNKKQVFSKIFYSDESGKTKWKGEGSDRGCCSEPVALEWEGKL 241
QY      261 IINTRVDYRRRLVYESSDMGNSWVEAVGTLSRVWGPSKSNQPSGSSSFTAVTIEGMRVM 320
Db      242 IINTRVDYRRRLVYESSDMGNSWLEAVGTLSRVWGPSKSNQPSGSSSFTAVTIEGMRVM 301
QY      321 LFTHPLNFKGRWLRDLNLWLTQRIYVNVGVQVSIQDENSAYSSVLYKDDKLYCLHEINS 380
Db      302 LFTHPLNFKGRWLRDLNLWLTQRIYVNVGVQVSIQDENSAYSSVLYKDDKLYCLHEINS 361
QY      381 NEYVSLVFARLVGELRIIKSVLQSWKQWDSHLSICTPADPAASSSERCSPAVTTVGLV 440
Db      362 NEYVSLVFARLVGELRIIKSVLQSWKQWDSHLSICTPADPAASSSERCSPAVTTVGLV 421
QY      441 GFLSHSATKTEWEDAYRCVNASTANAEVNPGLKTFAGVGGGALWPVSQQQGNQRYHFANH 500
Db      422 GFLSHSATKTEWEDAYRCVNASTANAEVNPGLKTFAGVGGGALWPVSQQQGNQRYHFANH 481
QY      501 AFTLVASVTIHEVPSPVASPLLGASLDSSGGKKLLGLSYDEKHQWQPIYGSTPTVTGSGWE 560
Db      482 AFTLVASVTIHEVPSPVASPLLGASLDSSGGKKLLGLSYDEKHQWQPIYGSTPTVTGSGWE 541
QY      561 MGKRYHVVLTMANKIGSVYIDGEPLEGSGQTVVDPDTPDISHFYVGVGYGRSDMPTISHV 620
Db      542 MGKRYHVVLTMANKIGSVYIDGEPLEGSGQTVVDPDTPDISHFYVGVGYGRSDMPTISHV 601
QY      621 TVNNVLLYNRLNAEIRTLFLSQDLIGTEAHWGSSGSS 660
Db      602 TVNNVLLYNRLNAEIRTLFLSQDLIGTEAHWGSSSDTS 641

RESULT 5
ADC21501
ID      ADC21501 standard; protein; 666 AA.
AC      ADC21501;
XX
DT      18-DEC-2003 (first entry)
XX
DE      T. cruzi trans-sialidase, TS, clone 19Y.
XX
KW      Trans-sialidase; TS; neurotrophic peptide; interleukin-6;
KW      IL-6 secretion inducing peptide; neuron; glial cell; trophic support;
KW      ciliary neurotrophic factor; CNTF; leukaemia inhibitory factor; LIF;
KW      amyotrophic lateral sclerosis; Alzheimer's disease; Parkinson's disease;
KW      Huntington's disease; Chagas' disease; peripheral neuropathy; palsy;
KW      multiple sclerosis; stroke; brain trauma; spinal cord trauma;
KW      peripheral nerve trauma.
XX
OS      Trypanosoma cruzi.
XX
FH      Key Location/Qualifiers
FT      Peptide 1..78
FT      /note= "Signal peptide"
FT      Protein 79..666
FT      /note= "Mature TS"
XX
FN      US2002137667-A1.
XX
PD      26-SEP-2002.
XX
PF      20-DEC-2000; 2000US-00745008.
XX
PR      20-DEC-1999; 99US-0172881P.
XX
PA      (TUFT ) UNIV TUFTS.
XX
PI      Chuenkova M, Pereira MA;

```


Query Match		76.2%; Score 2681; DB 7; Length 1162;
Best Local Similarity		86.5%; Pred. No. 1.7e-218;
Matches		525; Conservative 22; Mismatches 50; Indels 10; Gaps 4;
QY	66	MVAIADARYETSDNLSLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIKGNK 125
Db	1	MVAIADARYETSDNLSLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIKGNK 60
QY	126	LYVLGVSYNSSRSYMTSHGDARDWIDLLAVGEVTKTAGGKITASIKWGPSVSLKEPFP 185
Db	61	LYVLGVSYSRSYSSHGDARDWIDLLAVGEVTKTAGGKITASIKWGPSVSLKEPFP 120
QY	186	EMEGMTNTOFLGAGVAIVASGNLVYPVQVTKKQVFSKIFYSEDEGKTWKGEGRSD 245
Db	121	EMEGMTNTOFLGAGVAIVASGNLVYPVQVTKKQVFSKIFYSEDDGKTWKGEGRSD 180
QY	246	FGCSEPVALEWEGKLIINTRVYRRRLVYESDMGNSWEAVGTLRSVMGPSKSNQPGS 305
Db	181	FGCSEPVALEWEGKLIINTRVDMKRLVYESDMKWPVEAVGTVSRVMGPSKSNQPGS 240
QY	306	QSSFTAVTIEGRVNLFTPLNFKGRWLDRLNLWLTQNRQIYVGVQVSGIDENSAYSSV 365
Db	241	QTSFTAVTIEGRVNLFTPLNFKGRCVDRNLWLTQNRQIYVGVQVSGIDENSAYSSV 300
QY	366	LYKODKLYCLHEINSNEVYSLVFLARLVGELRIIKSVLQSWKNWDSLSSICTPADPA 425
Db	301	LYKODKLYCLHEINTDEVYSLVFLARLVGELRIIKSVLRSWKNW---TATCPAPAPLIQ 356
QY	426	SERGGCPAVTVTG---LVGFLSHSATKTEWEDAYRCVNASTANAERVPNGLKFGAGV 481
Db	357	PLRQRVVVPLSRVLVLLAFRCORLPK-RMGGSYRCVNASTANAERVRNGLKFGAGV 415
QY	482	ALMPVQSQGNQRHFANHAFTLVASVTIHEVPVSPFLGASLDSSGGKLLGLSYDEK 541
Db	416	ALMPVQSQGNQRHFANHAFTLVASVTIHEAPRAASPLGASLDSSGGKLLGLSYDEK 475
QY	542	HOWOPIYGSTPTPTGSMWCKRVHVLTWANKIGSVYIDGEPLEGSGQTVPDGRTPDI 601
Db	476	HOWOPIYGSTPTPTGSMWCKRVHVLVTWANKIGSVYIDGELLEGSGQTVPDGRTPDI 535
QY	602	SHFYVGYGRSDMPTTISHVTNNVLLYN-RQLNAEIRTLFLSODLIGTEAHMGSSSGS 660
Db	536	SHFYVGYGRSDMPTTISHVTNNVLLYNRQLNTEIRTLFLSODLIGTEAHMDSDDSS 595
QY	661	ERSTPGS 667
Db	596	AHSTPST 602
RESULT 7		
ADC21529		
ID	ADC21529	standard; protein; 200 AA.
XX	XX	
AC	ADC21529;	
XX	XX	
DT	18-DEC-2003	(first entry)
XX	XX	
DE	DE	T. cruzi trans-sialidase, TS, catalytically active fragment TS 154.
KW	KW	Trans-sialidase; TS; neurotrophic peptide; interleukin-6;
KW	KW	IL-6 secretion inducing peptide; neuron; glial cell; trophic support;
KW	KW	ciliary neurotrophic factor; CNTF; leukaemia inhibitory factor; LIF;
KW	KW	amyotrophic lateral sclerosis; Alzheimer's disease; Parkinson's disease;
KW	KW	Huntington's disease; Chagas' disease; peripheral neuropathy; palsy;
KW	KW	multiple sclerosis; stroke; brain trauma; spinal cord trauma;
XX	XX	peripheral nerve trauma.
OS	OS	Trypanosoma cruzi.
XX	XX	
PN	PN	US2002137667-A1.
XX	XX	
PD	26-SEP-2002.	
Query Match		30.3%; Score 1067; DB 7; Length 200;
Best Local Similarity		100.0%; Pred. No. 2.7e-82;
Matches		200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	219	KKQVFSKIFYSEDEGKTWKGEGRSDFGCSEPVALEWEGKLIINTRVYRRRLVYESD 278
Db	1	KKQVFSKIFYSEDEGKTWKGEGRSDFGCSEPVALEWEGKLIINTRVYRRRLVYESD 60
QY	279	MGNSWEAVGTLRSVMGPSKSNQPGSQSFTAVTIEGRVNLFTPLNFKGWLDRNL 338
Db	61	MGNSWEAVGTLRSVMGPSKSNQPGSQSFTAVTIEGRVNLFTPLNFKGWLDRNL 120
QY	339	LWLTNQRINYVGVQVSGIDENSAYSVLYKDDKLYCLHEINSNEVYSLVFLARLVGELRII 398
Db	121	LWLTNQRINYVGVQVSGIDENSAYSVLYKDDKLYCLHEINSNEVYSLVFLARLVGELRII 180
QY	399	KSVLQSKWNSDLSLSICTP 418
Db	181	KSVLQSKWNSDLSLSICTP 200
RESULT 8		
ADC21530		
ID	ADC21530	standard; protein; 200 AA.
XX	XX	
AC	ADC21530;	
XX	XX	
DT	18-DEC-2003	(first entry)
XX	XX	
DE	DE	T. cruzi trans-sialidase, TS, catalytically inactive fragment TS H32.
KW	KW	Trans-sialidase; TS; neurotrophic peptide; interleukin-6;

KW IL-6 secretion inducing peptide; neuron; glial cell; trophic support;
KW ciliary neurotrophic factor; CNTF; leukaemia inhibitory factor; LIF;
KW amyotrophic lateral sclerosis; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; Chagas' disease; peripheral neuropathy; palsy;
KW multiple sclerosis; stroke; brain trauma; spinal cord trauma;
KW peripheral nerve trauma; mutant; mutein.
XX
OS Synthetic.
OS Trypanosoma cruzi.
XX
XX US2002137667-A1.
XX
XX 26-SEP-2002.
PD
XX
XX 20-DEC-2000; 2000US-00745008.
PF
XX
XX 20-DEC-1999; 99US-0172881P.
PR
XX
XX (TUFT) UNIV TUFTS.
XX
XX Chuenkova M, Persira MA;
PI
XX
XX WPI; 2003-786654/74.
DR
XX
XX T. cruzi trans-sialidase derived neurotrophic and interleukin-6 secretion
PT inducing peptides, useful in the treatment of neuronal degeneration
PT caused for example by Chagas' disease.
PT
XX
XX Example 3; SEQ ID NO 31; 79pp; English.
PS
XX
XX The invention relates to a T. cruzi trans-sialidase (TS) derived
CC neurotrophic peptide appearing as ADC21513 or ADC21511 called C44 and
CC C14, or its variant, and an interleukin-6 (IL-6) secretion inducing
CC peptide appearing as ADC21511 called TR-1 (terminal repeat 1). Also
CC included are a composition comprising the peptides, fusion protein
CC comprising the peptides and a fusion partner, a composition comprising
CC the fusion protein and a physiological acceptable carrier, providing
CC trophic support for neurons or glial cells in a mammal (comprising
CC administering a therapeutically effective amount of T. cruzi trans-
CC sialidase (TS) or a neurotrophic variant) and stimulating (M3) the
CC secretion of IL-6 (interleukin-6) in a mammal comprising administering TS
CC or an IL-6 inducing variant. The fusion partner comprises a mammalian
CC neurotrophic factor which is ciliary neurotrophic factor (CNTF) or
CC leukaemia inhibitory factor (LIF). The peptides are useful in providing
CC trophic support for neurons and glial cells in a mammal suffering a
CC condition selected from: amyotrophic lateral sclerosis, Alzheimer's
CC disease, Parkinson's disease, Huntington's disease, Chagas' disease,
CC peripheral neuropathy, palsies, multiple sclerosis, stroke, brain trauma,
CC spinal cord trauma and peripheral nerve trauma, and in stimulating the
CC secretion of IL-6. The present sequence represents the trans-sialidase
CC catalytically inactive fragment TS H32 (has 6 substitutions in the
CC catalytic domain).
XX
XX Sequence 200 AA;
SQ

Query Match 29.5%; Score 1037; DB 7; Length 200;
Best Local Similarity 97.0%; Pred. No. 9.6e-80;
Matches 194; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
QY 219 KKKQVPSKIPYSEDEKTKWFGGRSDPGCEBPVALEWEGKLIINTRVDYRRRLVYESSD 278
Db 1 KKKQVPSKIPYSEDEKTKWFGGRSAFGCEAVALWEGKLIINTRVDYRRRLVYESSD 60
QY 279 MGNWVEAVGTLSRVWGPSPKSNQPSQSFTAVTIEGMRVMLFTHPLNFKGRWLRDLN 338
Db 61 MGNWTLEAVGTLSRVWGPSPKSNQPSQSFTAVTIEGMRVMLFTHPLNFKGRWLRDLN 120
QY 339 LMLTDNQRINYGVQSIGDENSAYSSVLKDDKLYCLCHEINSNEVYSLVFARLVGELRII 398
Db 121 LMLTDNQRINYGVQSIGDENSASHSVLYKDDKLYCLCHEINSNEVYSLVFARLVGELRII 180
QY 399 KSVLQSKWKNWDSHLSICTP 418
|||||

Db 181 KSVLQSKWKNWDSHLSICTP 200
RESULT 9
AAY44454
ID AAY44454 standard; protein; 1003 AA.
XX
XX AAY44454;
XX
XX 27-MAR-2000 (first entry)
XX
XX T. cruzi complement regulating protein.
XX
XX T. cruzi complement regulatory protein; CRP; vaccine;
KW Trypanosoma cruzi strain Y; Chagas' disease;
KW recombinant CRP eukaryotic expression cassette.
XX
XX Trypanosoma cruzi.
OS
XX W09960130-A1.
XX
XX 25-NOV-1999.
XX
XX 18-MAY-1999; 99WO-US010977.
PF
XX
XX 21-MAY-1998; 98US-0086197P.
PR
XX
XX (UYPI-) UNIV PITTSBURGH.
XX
XX Norris KA;
PI
XX
XX WPI; 2000-116315/10.
DR
XX N-PSDB; AA229719.
XX
XX A new vector encoding Trypanosoma cruzi complement regulatory protein,
PT for treatment of Chaga's disease.
XX
XX Example 1; Page 21-25; 33pp; English.
XX
XX The present sequence is Trypanosoma cruzi complement regulatory protein.
CC This is encoded by a full length coding sequence from strain Y of T.
CC cruzi. This sequence is used to construct a recombinant T. cruzi CRP
CC eukaryotic expression cassette. Recombinant CRP produced from host cells
CC can be used as a vaccine to prime the immune system of an animal.
CC Hybridomas secreting monoclonal antibodies recognising CRP are produced.
CC This is used to detect Chagas's disease-related proteins and for
CC treatment of the disease
XX
XX Sequence 1003 AA;
SQ

Query Match 25.7%; Score 903; DB 3; Length 1003;
Best Local Similarity 34.9%; Pred. No. 3e-67;
Matches 238; Conservative 88; Mismatches 270; Indels 85; Gaps 23;
QY 53 SFRLPALNVVDGVMVAIADARYETSDN-SLIDTVAKYSVDDGETWETQAIKNRSAS-S 110
Db 58 SFRAPSLAYVGVVATVEAHYTNSTDNKSVCVLAARSMESSGGGTNGTIVFDHYDK 117
QY 111 VSRVVDPTIV--KGNKLYVLVGSYNSRSRYWTS-HGDARDWDILLAVGEVTKSTAGGI 167
Db 118 IDRLSPFTFVDERDGNATNALVGGYGTSTPTLTVTDGKYWAPRIAAGSLIPYDDEEK- 176
QY 168 TASIKWGSVPVLSKEFFPAEMEGMHTN-----OPLGGAGVAI-VASNGNLVYPVQVTKKK 221
Db 177 -KEPKNQVASTSGVPHDLWESERTNPKPKQFLGGGACIKWEDDGRYVLPQALKDDG 235
QY 222 QVFSKIFYSESGTKTWKFGESRDFGCEBPVALEW-EGKLIINTRVDYRRRLVYESSDMG 280
Db 236 KVVSLVILAKKTSYGWEFSGNGTSDGCIQPAVLEWKEKELIMMTSCDDGSRVYRSSTMG 295
QY 281 NSWVEAVGTLSRVWGPSPKSNQPSQSFTAVTIEGMRVMLFTHPL-NFKGRWLRDLN 339
Db 296 NLWTEYDTLSRVWGNSTRVGVGHGAQGGFVSAMIDGQKVLVSRPVYSKDKETGRLLH 355

XX Trypanosoma cruzi.
XX Key Location/Qualifiers
FH Misc-difference 149
FT /note= "corresponds to CTA codon"
FT

XX WO9318787-Al.
XX 30-SEP-1993.
XX 25-MAR-1993; 93WO-US002869.
XX 25-MAR-1992; 92US-00857519.
XX 10-NOV-1992; 92US-00973851.
XX (UYN Y) UNIV NEW YORK STATE.
XX Nussenzweig V, Schenkman S, Eichinger D, Vandekerckhove F;
XX WPI; 1993-320452/40.
XX N-PSDB; AAQ49597.

XX New trans-sialidase polypeptide(s) isolated from Trypanosoma - used for
FT transferring sialic acid or for treating or preventing trypanomastigote
FT infection.
XX Disclosure; Fig 23; 130pp; English.

XX Clones 121 and 151, isolated from T.cruzi DNA library by homology to
CC known neuramidase sequences, were found to have identical sequences in
CC the region of the trans-sialidase gene necessary for enzymatic activity.
CC The 121/151-encoded amino acid sequence (AAR42015) is similar to the
CC known TCNA sequence (AAR42016) but distinct from the sequence encoded by
CC clone 154 (AAR42017). (Updated on 25-MAR-2003 to correct PN field.)
XX Sequence 166 AA;

Query Match 25.1%; Score 884; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 7.6e-67;
Matches 166; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 205 ASNGNLVYPVQVNTKKQVFSKIFYSEDEGKTWKFGESRDFGCSBPVALEWEGKLIINT 264
Db 1 ASNGNLVYPVQVNTKKQVFSKIFYSEDEGKTWKFGESRDFGCSBPVALEWEGKLIINT 60
QY 265 RVDYRRRLVYESSDMGNSWVEAVGTLRSVWGSPSPKSNQPGSQSSFTAVTIEGMRVLMFT 324
Db 61 RVDYRRRLVYESSDMGNSWVEAVGTLRSVWGSPSPKSNQPGSQSSFTAVTIEGMRVLMFT 120
QY 325 PLNFKGRWLRDRNLNMLTDNQRIYNYGVQVSGIDENSAYSSVLYKDD 370
Db 121 PLNFKGRWLRDRNLNMLTDNQRIYNYGVQVSGIDENSAYSSVLYKDD 166

RESULT 12
AAR42014
ID AAR42014 standard; protein; 166 AA.
XX AAR42014;
XX 25-MAR-2003 (revised)
DT 26-APR-1994 (first entry)
XX Protein with trans-sialidase and/or neuramidase activity.
XX Trypanosome; trans-sialidase; neuramidase; sleeping sickness;
KW Chagas' Disease; parasite.
XX Trypanosoma cruzi.
XX Key Location/Qualifiers
FH Misc-difference 149

/note= "corresponds to CTA codon"

FT WO9318787-Al.
XX 30-SEP-1993.
XX 25-MAR-1993; 93WO-US002869.
XX 25-MAR-1992; 92US-00857519.
XX 10-NOV-1992; 92US-00973851.
XX (UYN Y) UNIV NEW YORK STATE.
XX Nussenzweig V, Schenkman S, Eichinger D, Vandekerckhove F;
XX WPI; 1993-320452/40.
XX N-PSDB; AAQ49594.

XX New trans-sialidase polypeptide(s) isolated from Trypanosoma - used for
FT transferring sialic acid or for treating or preventing trypanomastigote
FT infection.
XX Claim 15; Fig 18; 130pp; English.
XX This is the amino acid sequence of the portion of trans-sialidase which
CC imparts trans-sialidase and/or neuramidase activity. (Updated on 25-MAR-
CC 2003 to correct PN field.)
XX Sequence 166 AA;

Query Match 25.1%; Score 884; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 7.6e-67;
Matches 166; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 205 ASNGNLVYPVQVNTKKQVFSKIFYSEDEGKTWKFGESRDFGCSBPVALEWEGKLIINT 264
Db 1 ASNGNLVYPVQVNTKKQVFSKIFYSEDEGKTWKFGESRDFGCSBPVALEWEGKLIINT 60
QY 265 RVDYRRRLVYESSDMGNSWVEAVGTLRSVWGSPSPKSNQPGSQSSFTAVTIEGMRVLMFT 324
Db 61 RVDYRRRLVYESSDMGNSWVEAVGTLRSVWGSPSPKSNQPGSQSSFTAVTIEGMRVLMFT 120
QY 325 PLNFKGRWLRDRNLNMLTDNQRIYNYGVQVSGIDENSAYSSVLYKDD 370
Db 121 PLNFKGRWLRDRNLNMLTDNQRIYNYGVQVSGIDENSAYSSVLYKDD 166

RESULT 13
AAR42015
ID AAR42015 standard; protein; 166 AA.
XX AAR42015;
XX 25-MAR-2003 (revised)
DT 26-APR-1994 (first entry)
XX Trans-sialidase/neuramidase encoded by clone 121/151.
XX Trypanosome; trans-sialidase; neuramidase; sleeping sickness;
KW Chagas' Disease; parasite.
XX Trypanosoma cruzi.
XX Key Location/Qualifiers
FH Misc-difference 149
FT /note= "corresponds to CTA codon"

XX WO9318787-Al.
XX 30-SEP-1993.
XX 25-MAR-1993; 93WO-US002869.
XX

PR 25-MAR-1992; 92US-00857519.
PR 10-NOV-1992; 92US-00973851.
XX (UUNY) UNIV NEW YORK STATE.
XX
XX
PI Nussenzweig V, Schenkman S, Eichinger D, Vandekerckhove F;
XX
XX WPI; 1993-320452/40.
XX N-PSDB; AAQ49595.
XX
XX New trans-sialidase polypeptide(s) isolated from Trypanosoma - used for
PT transferring sialic acid or for treating or preventing trypanomastigote
PT infection.
XX
XX Claim 22; Fig 23; 130pp; English.
XX
XX Clones 121 and 151, isolated from T.cruzi DNA library by homology to
CC known neuramidase sequences, were found to have identical sequences in
CC the region of the trans-sialidase gene necessary for enzymatic activity.
CC The 121/151-encoded amino acid sequence (AAR42015) is similar to the
CC known TCNA sequence (AAR42016) but distinct from the sequence encoded by
CC clone 154 (AAR42017). (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 166 AA;
SQ
Query Match 24.3%; Score 854; DB 2; Length 166;
Best Local Similarity 96.4%; Pred. No. 2.7e-64;
Matches 160; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
QY 205 ASNGNLVYPVQVTKKKQVFSKIPYSEDEGKTKWFGGRSDFGCSEPVALEWEGKLIINT 264
Db 1 ASNGNLVYPVQVTKKKQVFSKIPYSEDEGKTKWFGGRSDFGCSEPVALEWEGKLIINT 60
QY 265 RVDYRRRLVYESSDMGNSWVEAVGTLRSRVGSPKSNQPSQSFTAVTIEGMRVMLFTH 324
Db 61 RVDYRRRLVYESSDMGNTWLEAVGTLRSRVGSPKSNQPSQSFTAVTIEGMRVMLFTH 120
QY 325 PLNFKGRLDRRLNLWLTNQRINYVGVQVSGIDENSAYSSVLYKDD 370
Db 121 PLNFKGRLDRRLNLWLTNQRINYVGVQVSGIDENSAHSSVLYKDD 166
RESULT 14
AAR42016
ID AAR42016 standard; protein; 166 AA.
XX
AC AAR42016;
XX
DT 25-MAR-2003 (revised)
DT 26-APR-1994 (first entry)
XX
DE TCNA Trans-sialidase/neuramidase.
XX
XX Trypanosome; trans-sialidase; neuramidase; sleeping sickness;
KW Chagas' Disease; parasite.
XX
OS Trypanosoma cruzi.
XX
FH Key Location/Qualifiers
FT Misc-difference 149
FT /note= "corresponds to CTA codon"
XX
PN W09318787-A1.
XX
PD 30-SEP-1993.
XX
XX 25-MAR-1993; 93WO-US002869.
XX
PR 25-MAR-1992; 92US-00857519.
PR 10-NOV-1992; 92US-00973851.
XX
XX (UUNY) UNIV NEW YORK STATE.
XX

PI Nussenzweig V, Schenkman S, Eichinger D, Vandekerckhove F;
XX
XX WPI; 1993-320452/40.
XX N-PSDB; AAQ49596.
XX
XX New trans-sialidase polypeptide(s) isolated from Trypanosoma - used for
PT transferring sialic acid or for treating or preventing trypanomastigote
PT infection.
XX
XX Claim 22; Fig 23; 130pp; English.
XX
XX Clones 121 and 151, isolated from T.cruzi DNA library by homology to
CC known neuramidase sequences, were found to have identical sequences in
CC the region of the trans-sialidase gene necessary for enzymatic activity.
CC The 121/151-encoded amino acid sequence (AAR42015) is similar to the
CC known TCNA sequence (AAR42016) but distinct from the sequence encoded by
CC clone 154 (AAR42017). (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 166 AA;
SQ
Query Match 23.5%; Score 825; DB 2; Length 166;
Best Local Similarity 92.8%; Pred. No. 8e-62;
Matches 154; Conservative 8; Mismatches 4; Indels 0; Gaps 0;
QY 205 ASNGNLVYPVQVTKKKQVFSKIPYSEDEGKTKWFGGRSDFGCSEPVALEWEGKLIINT 264
Db 1 ASNGNLVYPVQVTKKKQVFSKIPYSEDEGKTKWFGGRSDFGCSEPVALEWEGKLIINT 60
QY 265 RVDYRRRLVYESSDMGNSWVEAVGTLRSRVGSPKSNQPSQSFTAVTIEGMRVMLFTH 324
Db 61 RVDYRRRLVYESSDMGKPEVAVGTVSRVWGFSPKSNQPSQSFTAVTIEGMRVMLFTH 120
QY 325 PLNFKGRLDRRLNLWLTNQRINYVGVQVSGIDENSAYSSVLYKDD 370
Db 121 PLNFKGRLDRRLNLWLTNQRINYVGVQVSGIDENSAYSSVLYKDD 166
RESULT 15
AAR26541
ID AAR26541 standard; protein; 618 AA.
XX
AC AAR26541;
XX
DT 17-OCT-2003 (revised)
DT 12-JAN-1998 (first entry)
XX
XX Trypanosoma cruzi antigen.
DE
XX
KW Antigen; epitope; vaccine; protective immunity; Chagas disease;
KW diagnosis; therapy; immunoassay.
XX
OS Trypanosoma cruzi; Tulahean strain C2.
XX
XX W09718475-A1.
PN
XX
PD 22-MAY-1997.
XX
XX 14-NOV-1996; 96WO-US018624.
XX
PR 14-NOV-1995; 95US-00557309.
XX
XX (CORI-) CORIXA CORP.
XX
XX Reed SG, Skeiky YAW, Lodes MJ, Houghton RL;
PI
XX
XX WPI; 1997-289413/26.
XX N-PSDB; AAT69167.
XX
XX Diagnosing Trypanosoma cruzi infection by detecting antibodies to novel
PT antigens - which are useful in vaccines to provide protective immunity
PT against Chagas' disease.
XX
XX Disclosure; Page 88-91; 110pp; English.
XX

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 8, 2004, 18:31:39 ; Search time 7675 Seconds
(without alignments)
7820.584 Million cell updates/sec

Title: US-10-086-913-1

Perfect score: 2010

Sequence: 1 atggcgagcagcatcatca.....cgcccgatccggtgctaa 2010

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthma:*
- 3: em_estin:*
- 4: em_estm:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_hic:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_hic:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: em_gss_hum:*
- 18: em_gss_inv:*
- 19: em_gss_pln:*
- 20: em_gss_vrt:*
- 21: em_gss_fun:*
- 22: em_gss_mam:*
- 23: em_gss_mus:*
- 24: em_gss_pro:*
- 25: em_gss_rod:*
- 26: em_gss_pbg:*
- 27: em_gss_vrl:*
- 28: gb_gss1:*
- 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	590.6	29.4	629	28	BH845080
2	584	29.1	672	28	BH841269
3	553.4	27.5	599	28	BH844958
4	551.2	27.4	652	28	BH844398

5	548.6	27.3	675	28	BH844084
6	500.8	24.9	532	28	BH842621
7	485.4	24.1	495	28	BH197094
8	460	22.9	526	28	BH844627
9	378.2	18.8	467	28	BH843788
10	378.2	18.8	467	28	BH843853
11	372	18.5	412	28	AQ911131
12	360.8	18.0	471	28	BH191346
13	342	17.0	366	28	AQ910653
14	330.6	16.4	367	28	BH844382
15	315.2	15.7	364	28	AQ911046
16	312	15.5	360	28	AQ906250
17	300	14.9	357	28	AQ907846
18	287.4	14.3	398	28	AQ902499
19	285.2	14.2	390	9	AA556106
20	279.6	13.9	304	28	AQ910261
21	266.6	13.3	689	28	BH842222
22	266.6	13.3	689	28	BH843018
23	214	10.6	305	28	BH197012
24	213	10.6	319	28	BH196504
25	210.6	10.5	257	28	BH194538
26	205.8	10.2	476	28	BH199280
27	189.2	9.4	236	28	BH192025
28	185.2	9.2	607	28	AQ953456
29	183.8	9.1	203	28	BH194952
30	175.6	8.7	215	28	BH192734
31	162.4	8.1	681	28	BH842371
32	162.4	8.1	681	28	BH844940
33	160	8.0	675	28	AQ953264
34	158.6	7.9	198	28	BH193191
35	155	7.7	699	28	AQ953610
36	141.6	7.0	665	29	TA155F11Q
37	141.2	7.0	623	28	AQ940568
38	139.4	6.9	331	28	BH191354
39	139	6.9	612	28	A2218427
40	138.6	6.9	609	28	BH846083
41	136.4	6.8	411	28	AQ444507
42	135.2	6.7	565	28	AQ951915
43	134.6	6.7	564	29	TA19G08P
44	134.4	6.7	665	28	AQ946989
45	133.8	6.7	708	28	BH844188

ALIGNMENTS

RESULT 1
BH845080
LOCUS TC3-53G11.TP TC3 Trypanosoma cruzi genomic clone TC3-53G11, genomic survey sequence.
DEFINITION
ACCESSION BH845080.1 GI:21415255
VERSION BH845080
KEYWORDS GSS.
SOURCE Trypanosoma cruzi
ORGANISM Trypanosoma cruzi
REFERENCE 1 (bases 1 to 629)
AUTHORS Myler, P.J., Aggarwal, G., Fazelina, G., Mack, J., Marty, A., Mullen, H., Nelson, S., Pentony, M., Rinta, J., Robertson, L., Seyler, A., Siek, E., Stuart, K., Vogt, C., Worthey, E., El-Sayed, N.M., Ghedin, B. and Andersson, B.
TITLE Trypanosoma cruzi CL-Brener TC3 BAC-end sequencing
JOURNAL Trypanosoma; Schizotrypanum.
COMMENT Other_GSSs: TC3-53G11.TV
Contact: Peter Myler
Seattle Biomedical Research Institute
4 Nickerson Street, Seattle, WA 98109, USA
Tel: 206 284 8846
Fax: 206 284 0313
Email: mylerp@sbri.org
Clones are derived from the Trypanosoma cruzi CL-Brener BAC library

TC3. For clone availability, please contact Dr. Bjorn Andersson at Uppsala University (bjorn.andersson@genpat.uu.se).
Seq primer: S86
Class: BAC ends.

FEATURES

source
1. .629
Location/Qualifiers
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/strain="CL Brener"
/db_xref="taxon:5693"
/clone="TC3-53G11"
/clone_lib="TC3"
/note="Vector: pBelobAC11; Site 1: Hin dIII; Constructed for Uppsala University by Marie-Christine Le Paslier in the laboratory of Denis Le Paslier at the Centre d'Etude du Polymorphisme Humain (CEPH), Paris, France. Briefly, Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained from Dr. Franco da Silveira) was partially digested with Hin dIII. High molecular weight fragments were ligated in pBelobAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 33 X the haploid genome."

ORIGIN

Query Match 29.4%; Score 590.6; DB 28; Length 629;
Best Local Similarity 98.3%; Pred. No. 3.1e-149;
Matches 618; Conservative 0; Mismatches 9; Indels 2; Gaps 2;
QY 449 ATATTCCTGTCGCGTGGTGAAGTCAAGATCACTGCGG-CGCGCAAGATAACTGGG 507
Db |||||
1 ATATTCCTGTCGCGTGGTGAAGTCAAGATCACTGCGGCGCGCAAGATAACTGGG 60
QY 508 AGTATCAAAATGGGGAGCCCGTGTCACTGAAGGAATTTTCCCGCGGAAATGGAGGA 567
Db |||||
61 AGTATCAAAATGGGGAGCCCGTGTCACTGAAGGAATTTTCCCGCGGAAATGGAGGA 120
QY 568 ATGCACACAATCAATTTCTTGGCGGTGCGAGGTGTTGCCATGTGGCGTCCACCGGGAAT 627
Db |||||
121 ATGCACACAATCAATTTCTTGGCGGTGCGAGGTGTTGCCATGTGGCGTCCACCGGGAAT 180
QY 628 CTCTGTACCTGTGACAGGTACGACAAAGCAAGCAAGTATTTTCCAGATCTCTAC 587
Db |||||
181 CTCTGTACCTGTGACAGGTACGACAAAGCAAGTATTTTCCAGATCTCTAC 240
QY 688 TCGAAGACGAGGCGCAAGACGTGGAAGTTTGGGAGGGTAGGAGTATTTGGCTGCTCT 747
Db |||||
241 TCGTATGATGATGCAAGACGTGGAAGTTTGGGAGGGTAGGAGTATTTGGCTGCTCT 300
QY 748 GAACCTGTGCGCTTGAGTGGGAGGGAGCTCATATAAACAATCGAGTTGACTATCGC 807
Db |||||
301 GAACCTGTGCGCTTGAGTGGGAGGGAGCTCATATAAACAATCGAGTTGACTATCGC 360
QY 808 CGCGCTGTGTTACAGTCACTGACATGGGGAATTCGTGGTGGAGGCTGTGCGCAGG 867
Db |||||
361 CGCGCTGTGTTACAGTCACTGACATGGGGAATTCGTGGTGGAGGCTGTGCGCAGG 420
QY 868 CTCTCAGTGTGTTGGGGCCCTTCCAAAATCGAACCCCGGAGTCAGAGCAGCTTC 927
Db |||||
421 CTCTCAGTGTGTTGGGGCCCTTCCAAAATCGAACCCCGGAGTCAGAGCAGCTTC 480
QY 928 ACTGCGCTGACCATTCAGGAGGATCGGTGTATGCTTTCACACCCCGCTGAATTTAAG 987
Db |||||
481 ACTGCGCTGACCATTCAGGAGGATCGGTGTATGCTTTCACACCCCGCTGAATTTAAG 540
QY 988 GGAAGTGTGTCGCGACCGACTGAACCTCTGGCTGACGATAACCGAGCGAATTTATAAC 1047
Db |||||
541 GGAAGTGTGTCGCGACCGACTGAACCTCTGGCTGACGATAACCGAGCGAATTTATAAC 600
QY 1048 GTT-GGGCAAGTATCCATTTGGTATGAAA 1075
Db |||||
601 GTTGGGCAAGTATCCATTTGGTATGAAA 629

RESULT 2

BH841269
LOCUS

DEFINITION

BH841269 672 bp DNA linear GSS 13-JUN-2002
TC3-56E8-TV TC3 Trypanosoma cruzi genomic clone TC3-56E8, genomic survey sequence.

ACCESSION

BH841269

BH841269.1 GI:21408484

VERSION

GSS.

KEYWORDS

Trypanosoma cruzi

SOURCE

Trypanosoma cruzi

ORGANISM

Trypanosoma, Schizotrypanum.

REFERENCE

1 (bases 1 to 672)

AUTHORS

Myler, P.J., Aggarwal, G., Fazelinia, G., Mack, J., Marty, A., Munden, H., Nelson, S., Pentony, M., Rinta, J., Robertson, L., Seyler, A., Sisk, E., Stuart, K., Vogt, C., Worthey, E., El-Sayed, N.M., Ghedin, E. and Andersson, B.

TITLE

Trypanosoma cruzi CL-Brener TC3 BAC-end sequencing

JOURNAL

Unpublished (2001)

COMMENT

Other_GSSs: TC3-56E8.TP

Contact: Peter Myler

Seattle Biomedical Research Institute

4 Nickerson Street, Seattle, WA 98109, USA

Tel: 206 284 8845

Fax: 206 284 0313

Email: mylerp@sbri.org

Clones are derived from the Trypanosoma cruzi CL-Brener BAC library

TC3. For clone availability, please contact Dr. Bjorn Andersson at

Uppsala University (bjorn.andersson@genpat.uu.se).

Seq primer: T7

Class: BAC ends.

FEATURES

Location/Qualifiers

1. .672

/organism="Trypanosoma cruzi"

/mol_type="genomic DNA"

/strain="CL Brener"

/db_xref="taxon:5693"

/clone="TC3-56E8"

/clone_lib="TC3"

/note="Vector: pBelobAC11; Site 1: Hin dIII; Constructed for Uppsala University by Marie-Christine Le Paslier in the laboratory of Denis Le Paslier at the Centre d'Etude du Polymorphisme Humain (CEPH), Paris, France. Briefly, Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained from Dr. Franco da Silveira) was partially digested with Hin dIII. High molecular weight fragments were ligated in pBelobAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 33 X the haploid genome."

ORIGIN

Query Match 29.1%; Score 584; DB 28; Length 672;
Best Local Similarity 91.8%; Pred. No. 2e-147;
Matches 617; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 374 AGCTTTACGTCCTGGTTGGAGCTACAAAGCTTCGAGGAGTCCAGCTCCGATGGTG 433
Db |||||
1 AGCTTTACGTTCTGGTTGGAGGATCAATATATCACTAATCTGATGTGGATGGTG 60
QY 434 ATGCGAGAGACTGGGATATTCTCTTCCCGTTGGTGGAGTCCAGAAAGTCCACTGCGGGCG 493
Db |||||
61 ATGCGAGGAGTGGGATATTCTCTTCCCGTTGGTGGAGTCCAGAAAGTCCACTGCGGGCG 120
QY 494 GCAAGATAACTGCGAGTATCAAAATGGGAGCCCGCTGTCTGACGAGTCCAGAGGAATTTTCCCGG 553
Db |||||
121 GCAATACACTGTGAGTATCAAAATGGGAGCCCGCTGTCTGACGAGTCCAGAGGAATTTTCCCGT 180
QY 554 CGGAAATGGAAGGAATGCGACACAAATCAATTTCTTGGCGGTGAGGTGTTGCCATTTGGTG 613
Db |||||
181 GGGAAATGGAAGGAGTGGCCACAAAGCAATTTCTTGGCGGTGCGGTGTTGCCACTCTGG 240
QY 614 CGTCCACCGGAATCTTGTGTACCTCTGTGAGGTACGAACAAAGAGCAAGTATTTT 673
Db |||||

```
Db 241 CGTCCAAAGGAATCTTGTTACCTGTGCAGGTTCAGAACGTGAAGGGAACAAATTTTCT 300
Qy 674 CCAAGATCTTCTACTCGGAAGACGAGGCAAGACGTGGAAAGTTTGGGGAGGTTAGGAGTG 733
Db 301 CCAAGATCTTCTACTCGGAAGATGATGCAAGACGTGGAAAGTTTGGGAAGGGTAGGAGCG 360
Qy 734 ATTTGGCTGCTCTGAACCTGTGGCCCTTGAGTGGGAGGGGAAGCTCATATAAACACTC 793
Db 361 CTTTGGCTGCTCTGAACCTGTGGCCCTTGAGTGGGAGGGGAAGCTCATATAAACACTC 420
Qy 794 GAGTTGACTATCGCCCGCTGCTGCTGACAGTCCAGTGAACATGGGGAATTCGTGGGTGG 853
Db 421 GAGTTGACCGGCAAGCCGCTGCTGCTGACAGTCCAGTGAACATGGGGAATTCGTGGGTGG 480
Qy 854 AGCTGTGCGACAGCTCTCAAGTGTGTGGGGCCCTTCACCAAAATCGAACAGCCCGGCA 913
Db 481 AGCTGTGCGACAGCTCTCAAGTGTGTGGGGCCCTTCACCAAAATCGAACAGCCCGGCA 540
Qy 914 GTCAGAGCAGCTTCACGTCGCTGACCATCGAGGAATCGGTGTTATGCTTTCACACACC 973
Db 541 GTCAGAGCAGCTTCACGTCGCTGACCATCGAGGAATCGGTGTTATGCTTTCACACACC 600
Qy 974 CGCTGAATTTTAAAGGGAAGTGTGCTGCGACCGACTGAACCTCTGCTGACCGATAACC 1033
Db 601 CGCTGAATTTTAAAGGGAAGTGTGCTGCGACCGACTGAACCTCTGCTGACCGATAACC 660
Qy 1034 AGCGCATTTATA 1045
Db 661 AGCGCATTTATA 672
```

```
RESULT 3
BH844958
LOCUS TC3-53G7.TV TC3 Trypanosoma cruzi genomic clone TC3-53G7, genomic
DEFINITION survey sequence.
```

```
ACCESSION BH844958
VERSION BH844958.1 GI:21415010
KEYWORDS GSS.
SOURCE Trypanosoma cruzi
ORGANISM Trypanosoma cruzi
```

```
REFERENCE 1 (bases 1 to 599)
AUTHORS Myler,P.J., Aggarwal,G., Fazelinia,G., Mack,J., Marty,A.,
Munden,H., Nelson,S., Pentony,M., Rinta,J., Robertson,L.,
Seyler,A., Sisk,E., Stuart,K., Vogt,C., Worthey,E., El-Sayed,N.M.,
Ghedin,E. and Andersson,B.
```

```
TITLE Trypanosoma cruzi CL-Brener TC3 BAC-end sequencing
JOURNAL Unpublished (2001)
COMMENT Other GSSs: TC3-53G7.TP
```

```
Contact: Peter Myler
Seattle Biomedical Research Institute
4 Nickerson Street, Seattle, WA 98109, USA
Tel: 206 284 8846
Fax: 206 284 0313
Email: mylerp@sbri.org
Clones are derived from the Trypanosoma cruzi CL-Brener BAC library
TC3. For clone availability, please contact Dr. Bjorn Andersson at
Uppsala University (bjorn.andersson@genpat.uu.se).
```

```
Seq primer: T7
Class: BAC ends.
```

```
FEATURES
source Location/Qualifiers
1..599
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/strain="CL Brener"
/db_xref="taxon:5693"
/clone="TC3-53G7"
/clone_lib="TC3"
/notes="Vector: pBelobAC11; Site_1: Hin dIII; Constructed
for Uppsala University by Marie-Christine Le Paslier in
the laboratory of Denis Le Paslier at the Centre d'Etude
```

du Polymorphisme Humain (CEPH), Paris, France. Briefly, Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained from Dr. Franco da Silveira) was partially digested with Hin dIII. High molecular weight fragments were ligated in pBelobAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 33 X the haploid genome."

ORIGIN

```
Query Match 27.5%; Score 553.4; DB 28; Length 599;
Best Local Similarity 96.2%; Pred. No. 4.1e-139;
Matches 578; Conservative 0; Mismatches 21; Indels 2; Gaps 1;

Qy 418 TGAACCTGCGATGTCATGCGAGAGACTGGGATATTTCTGCTTCGCGTTGTGTAGGTCACG 477
Db 1 TGAACCTGCGATGTCATGTC--GATACCTGGGATATCTGCTTCGCGTTGTGTAGGTCACG 58
Qy 478 AGTTCACACTCGCGCGCGCAAGATAACTGCGAGTATCAAAATGGGGAGCCCGGTGTCACATG 537
Db 59 AGTTCACACTCGCGCGCGCAAGATAACTGCGAGTATCAAAATGGGGAGCCCGGTGTCACATG 118
Qy 538 AAGGAATTTTTCGCCGCGGAAATGGAAGGAATGCACACAAATCAATTTCTTGGCGGTGCA 597
Db 119 AAGGAATTTTTCGCCGCGGAAATGGAAGGAATGCACACAAATCAATTTCTTGGCGGTGCA 178
Qy 598 GGTGTTGCCATTTGTGCGCTCCAAACGGGAATCTTGTGTATCCCTGTGCAGGTTCAGAACAA 657
Db 179 GGTGTTGCCATTTGTGCGCTCCAAACGGGAATCTTGTGTATCCCTGTGTATTTTACGAACAA 238
Qy 658 AGAAGCAAGTTTTCCTCAAGATCTTCTACTCGGAAGACGGGCAAGAGCTGGAAGTTT 717
Db 239 AGAAGCAAGTTTTCCTCAAGATCTTCTACTCGGAAGATGATGGCAAGAGCTGGAAGTTT 298
Qy 718 GGGAGGGTAGGAGTGAATTTTGGCTGCTCTGTGAACCTGTGGCCCTTGAGTGGGAGGGAAG 777
Db 299 GGGAGGGTAGGAGTGAATTTTGGCTGCTCTGTGAACCTGTGGCCCTTGAGTGGGAGGGAAG 358
Qy 778 CTCATCATAAACACTCGAGTTGACTATCGCCCGCTGTGTTGATCGAGTCCAGTGCATG 837
Db 359 CTCATCATAAACACTCGAGTTGACTATCGCCCGCTGTGTTGATCGAGTCCAGTGCATG 418
Qy 838 GGGAAATTCGTGGTGAGGCTGTGCGACGCTCTCACTGCTGTGTGGGCCCTTCACCAAAA 897
Db 419 GGGAAATTCGTGGTGAGGCTGTGCGACGCTCTCACTGCTGTGTGGGCCCTTCACCAAAA 478
Qy 898 TCGAACACAGCCCGGAGTCAGAGCAGCTTCACCTGCGGTGACCATCGAGGGAATGCGTGT 957
Db 479 TCGAACACAGCCCGGAGTCAGAGCAGCTTCACCTGCGGTGACCATCGAGGGAATGCGTGT 538
Qy 958 ATGCTCTTCACACACCCCGCTGAATTTTAAAGGGAAGTGTGCTGCGACCGACTGAACCTC 1017
Db 539 ATGCTCTTCACACACCCCGCTGAATTTTAAAGGGAAGTGTGCTGCGACCGACTGAACCTC 598
Qy 1018 T 1018
Db 599 T 599
```

RESULT 4

```
BH844398
LOCUS TC3-56J15.TV TC3 Trypanosoma cruzi genomic clone TC3-56J15, genomic
DEFINITION survey sequence.
ACCESSION BH844398
VERSION BH844398.1 GI:21413883
KEYWORDS GSS.
SOURCE Trypanosoma cruzi
ORGANISM Trypanosoma cruzi
REFERENCE 1 (bases 1 to 652)
AUTHORS Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Myler,P.J., Aggarwal,G., Fazelinia,G., Mack,J., Marty,A.,
Munden,H., Nelson,S., Pentony,M., Rinta,J., Robertson,L.,
```

Seyler, A., Sisk, E., Stuart, K., Vogt, C., Worthey, E., El-Sayed, N.M., Ghedin, E. and Anderson, B.
Trypanosoma cruzi CL-Brener TC3 BAC-end sequencing
Unpublished (2001)
Other GSSs: TC3-56J15.TP
Contact: Peter Wyler
Seattle Biomedical Research Institute
4 Nickerson Street, Seattle, WA 98109, USA
Tel: 206 284 8846
Fax: 206 284 0313
Email: mylerpj@sbri.org
Clones are derived from the Trypanosoma cruzi CL-Brener BAC library
TC3. For clone availability, please contact Dr. Bjorn Andersson at
Uppsala University (bjorn.andersson@genpat.uu.se).
Seq primer: T7
Class: BAC ends.

FEATURES

source
1. .652
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/strain="CL Brener"
/db_xref="taxon:5693"
/clone="TC3-56J15"
/clone_lib="TC3"
/note="Vector: pBelOBAC11; Site 1: Hin dIII; Constructed for Uppsala University by Marie-Christine Le Paslier in the laboratory of Denis Le Paslier at the Centre d'Etude du Polymorphisme Humain (CEPH), Paris, France. Briefly, Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained from Dr. Franco da Silveira) was partially digested with Hin dIII. High molecular weight fragments were ligated in pBelOBAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 33 X the haploid genome."

ORIGIN

Query Match 27.4%; Score 551.2; DB 28; Length 652;
Best Local Similarity 90.3%; Pred. No. 1.7e-138;
Matches 589; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
QY 401 ACAGTTTCGAGGAGTACTGACGTCGCGATGTCATGCGAGAGTGGATATTCGCTTG 460
Db 1 ATATATCACTACTGATGTCATGTCGATGTCGAGTATCGGGATATTCGCTTG 60
QY 461 CCGTGTGTGAGGTACAGAGTCCACTGCGCGCGCAAGATACTGCGAGTATCAAAATGGG 520
Db 61 CCGTGTGTGAGGTACAGAGTCCACTGCGCGCGCAAGATACTGCGAGTATCAAAATGGG 120
QY 521 GGAGCCCGGTGTCACGTAAGGAATTTTCCCGCGGGAATGGAGGAATGCACAAATC 580
Db 121 GGAGCCCGGTGTCACGTAAGGAATTTTCCCGCGGGAATGGAGGAATGCACAAATC 180
QY 581 AATTTCTTGGCGGTGCGAGGTGTGCCATTTGCGGTCCAAAGGGAATCTGTGACCTGG 640
Db 181 AATTTCTTGGCGGTGCGAGGTGTGCCATTTGCGGTCCAAAGGGAATCTGTGACCTGG 240
QY 641 TGCAGGTACGAAACAAAAGAACAGTATTTTCCAAAGATCTTCTACTCGGAAGACGAG 700
Db 241 TGCAGGTACGAAACAAAAGAACAGTATTTTCCAAAGATCTTCTACTCGGAAGATGATG 300
QY 701 GCAGACGTGGAAGTTTGGGAGGGTAGGATGATTTTGGCTGCTCTGAACCTGTGGCCC 760
Db 301 GCAGACGTGGAAGTTTGGGAGGGTAGGATGATTTTGGCTGCTCTGAACCTGTGGCCC 360
QY 761 TTGAGTGGGAGGGGAAGCTCATATAACACTCGAGTTGACTATCGCCCGCTCTGGTGT 820
Db 361 TTGAGTGGGAGGGGAAGCTCATATAACACTCGAGTTGACTATCGCCGCAAGCCCTCTGGTGT 420
QY 821 ACAGTCCAGTGCATGCGGAATTCGTGGGTGAGAGCTGTGCGACGCTCTCAAGCTGTGT 880
Db 421 ACAGTCCAGTGCATGCGGAATTCGTGGGTGAGAGCTGTGCGACGCTCTCAAGCTGTGT 480
QY 881 GGGGCCCTCACCAAAATCGAACCCGCGGAGTCAGAGCAGCTTCACTGCGGTGACCA 940

Db 481 GGGGCCCTCACCAAAATCGAACCCGCGGAGTCAGAGCAGCTTCACTGCGGTGACCA 540
QY 941 TCGAGGGAATGCGTGTATGCTCTTCACACACCGCTGAATTTTAAGGGAAGTGGCTGC 1000
Db 541 TCGAGGGAATGCGTGTATGCTCTTCACACACCGCTGAATTTTAAGGGAAGTGGCTGC 600
QY 1001 GCGACCGACTGAACCTCTGCTGACGATAACGAGCGCATTTATACGTTGG 1052
Db 601 GCGACCGACTGAACCTCTGCTGACGATAACGAGCGCATTTATACGTTGG 652

RESULT 5

BH844084
LOCUS BH844084
DEFINITION BH844084
ACCESSION BH844084
VERSION BH844084.1 GI:214113249
KEYWORDS GSS.
SOURCE Trypanosoma cruzi
ORGANISM Trypanosoma cruzi
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
1 (bases 1 to 675)

REFERENCE

AUTHORS Myler, P.J., Aggarwal, G., Fazelinia, G., Mack, J., Marty, A., Munden, H., Nelson, S., Pentony, M., Rinta, J., Robertson, L., Seyler, A., Sisk, E., Stuart, K., Vogt, C., Worthey, E., El-Sayed, N.M., Ghedin, E. and Anderson, B.
Trypanosoma cruzi CL-Brener TC3 BAC-end sequencing
Unpublished (2001)
Other GSSs: TC3-53M19.TP

TITLE

Trypanosoma cruzi CL-Brener TC3 BAC-end sequencing
Unpublished (2001)
Other GSSs: TC3-53M19.TP
Contact: Peter Wyler
Seattle Biomedical Research Institute
4 Nickerson Street, Seattle, WA 98109, USA
Tel: 206 284 8846
Fax: 206 284 0313
Email: mylerpj@sbri.org
Clones are derived from the Trypanosoma cruzi CL-Brener BAC library
TC3. For clone availability, please contact Dr. Bjorn Andersson at
Uppsala University (bjorn.andersson@genpat.uu.se).
Seq primer: T7
Class: BAC ends.

COMMENT

Location/Qualifiers
1. .675
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/strain="CL Brener"
/db_xref="taxon:5693"
/clone="TC3-53M19"
/clone_lib="TC3"
/note="Vector: pBelOBAC11; Site 1: Hin dIII; Constructed for Uppsala University by Marie-Christine Le Paslier in the laboratory of Denis Le Paslier at the Centre d'Etude du Polymorphisme Humain (CEPH), Paris, France. Briefly, Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained from Dr. Franco da Silveira) was partially digested with Hin dIII. High molecular weight fragments were ligated in pBelOBAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 33 X the haploid genome."

FEATURES

source
1. .675
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/strain="CL Brener"
/db_xref="taxon:5693"
/clone="TC3-53M19"
/clone_lib="TC3"
/note="Vector: pBelOBAC11; Site 1: Hin dIII; Constructed for Uppsala University by Marie-Christine Le Paslier in the laboratory of Denis Le Paslier at the Centre d'Etude du Polymorphisme Humain (CEPH), Paris, France. Briefly, Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained from Dr. Franco da Silveira) was partially digested with Hin dIII. High molecular weight fragments were ligated in pBelOBAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 33 X the haploid genome."

ORIGIN

Query Match 27.3%; Score 548.6; DB 28; Length 675;
Best Local Similarity 88.3%; Pred. No. 8.8e-138;
Matches 596; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
QY 374 AGCTTTACGCTCGTGTGGAGCTACCAAGTTCGAGAGCTACTGACGCTGCGATGGT 433
Db 1 AGCTTTACGCTCGTGTGGAGCTACCAAGTTCGAGAGCTACTGACGCTGCGATGGT 60
QY 434 ATGCGAGAGCTGGGATATTCGCTGCGCTGGTGGAGTTCAGAGTCCACTGCGGCG 493

Db 61 ATGGAACGCACTGGGATATTTCTGCTGCCGTTGGTGAGGTTACGAAGTCCACTGCGAGACG 120
Qy 494 GCAAGATAACTGCGAGTATCAAAATGGGGAGCCCGGTGTCACTGAAGAAATTTTCCCGG 553
Db 121 GCAAGCAACTGCGAATATCACATGGGGAGTCCCGTGTCACTGAAGAAATTTTCCCGG 180
Qy 554 CGGAATGGAAGAAATGACACAAATTTCTTGGCGGTGACAGGTGTTGCCATTTGTGG 613
Db 181 CATACATGGAAGAAATACCTACAAAGCAATTTCTCGCGCGTGCAGGTGTTGCCACTGTGG 240
Qy 614 CGTCCAACGGGAATCTTGCTGACCTGTGACGGTTACGAACAAAGAAAGCAAGTTTCTT 673
Db 241 CGTCCAATGGGAATCTTGCTGACCTGTGACGGTTACGAACAAAGAAAGCAAGTTTCTT 300
Qy 674 CCAAGATCTTTCTACTCGGAAGACGAGGCAAGACGTGGAAGTTTGGGGAGGGTAGGATG 733
Db 301 CCAAGATCTTTTATTCGGCAGATGAGGGCAAGACGTGGAATTTCTCAGAGGGTAGGAGCG 360
Qy 734 ATTTTGGCTCTGTAACCTGTGGCCCTTGAGTGGGAGGGGAAGCTCATCATAAACACTC 793
Db 361 ATTTTGGCTCTCGGAACCTGTGGTCTTGAGTGGAAAAATAAATTCATCGTAAACACCC 420
Qy 794 GAGTTCACATATCGCCCGCTGCTGCTGACAGTCCAGTGCAGTGGGAATTCGTGGGTGG 853
Db 421 GGGTTGACCGGGCGCGCCGCTGCTGGTGACAAATTCGGTGACATGAGAGACCGTGGGTGG 480
Qy 854 AGCGTGTGCGCAGCGCTCTCAACGTGTGTGGGGCCCTCACCAAAATCGAACCGCCGGCA 913
Db 481 AGCGTGTGCGCAGCGCTCTCGCGTGTGTGGGGCCCTCACCAAAATCGAACCGCCGGCA 540
Qy 914 GTCAAGAGCTTCACTGCGCGTGACCAATCGAGGGAATGCGTGTATGCTTTTCACACACC 973
Db 541 GTCAAGAGCTTCACTGCGCGTGACCAATCGAGGGAATGCGTGTATGCTTTTCACACACC 600
Qy 974 CGCTGAATTTTAAGGGAAGTGGCTGCGCAGCGACTGACCTCGCTGACCGGATAACC 1033
Db 601 CGCTGAATTTTAAGGGAAGTGGCTGCGCAGCGACTGACCTTTGGCTGACCGGATAACC 660
Qy 1034 AGCGCATTTTATAACG 1048
Db 661 AGCGCATTTTATAACG 675

RESULT 6
BH842621
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BH842621 532 bp DNA linear GSS 13-JUN-2002
TC3-53123.TP TC3 Trypanosoma cruzi genomic clone TC3-53123, genomic survey sequence.
BH842621
BH842621.1 GI:21409836
GSS.
Trypanosoma cruzi
Trypanosoma cruzi
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
1 (bases 1 to 532)
Myler,P.J., Aggarwal,G., Fazelinia,G., Mack,J., Marty,A., Munden,H., Nelson,S., Pentony,M., Rinta,J., Robertson,L., Seyler,A., Sisk,E., Stuart,K., Vogt,C., Worthey,E., El-Sayed,N.M., Ghedin,E. and Andersson,B.
Trypanosoma cruzi CL-Brener
Unpublished (2001)
Other GSSs: TC3-53123.TV
Contact: Peter Myler
Seattle Biomedical Research Institute
4 Nickerson Street, Seattle, WA 98109, USA
Tel: 206 284 8846
Fax: 206 284 0313
Email: mylerp@bri.org
Clones are derived from the Trypanosoma cruzi CL-Brener BAC library TC3. For clone availability, please contact Dr. Bjorn Andersson at Uppsala University (bjorn.andersson@genpat.uu.se).
Seq primer: SP6

Class: BAC ends.
Location/Qualifiers
1. .532
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/strain="CL Brener"
/db_xref="taxon:5693"
/clone="TC3-53123"
/clone_lib="TC3"
/notes="Vector: pBeloBAC11; Site 1: Hin dIII; Constructed for Uppsala University by Marie-Christine Le Paslier in the laboratory of Denis Le Paslier at the Centre d'Etude du Polymorphisme Humain (CEPH), Paris, France. Briefly, Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained from Dr. Franco da Silveira) was partially digested with Hin dIII. High molecular weight fragments were ligated in pBeloBAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 33 X the haploid genome."

ORIGIN
Query Match 24.9%; Score 500.8; DB 28; Length 532;
Best Local Similarity 97.6%; Pred. No. 8e-125;
Matches 519; Conservative 0; Mismatches 12; Indels 1; Gaps 1;
Qy 497 AGATAACTCGAGTATCAAAATGGGGAGCCCGGTGTCACTGAAGAAATTTTCCCGCGG 556
Db 1 AGATAACTCGAGTATCAAAATGGGGAGCCCGGTGTCACTGAAGAAATTTTCCCGCGG 60
Qy 557 AAATGGAAGGAATGACACAAATCAATTTCTTGGCGGTGACAGGTGTTGCCATTTGTGGCGT 616
Db 61 AAATGGAAGGAATGACACAAATCAATTTCTTGGCGGTGACAGGTGTTGCCATTTGTGGCGT 120
Qy 617 CCAACCGGGAATCTTGTGTACCTGTGACGTTACGAACAAAGAAAGCAAGTTTTCCTCA 676
Db 121 CCAACCGGGAATCTTGTGTACCTGTGACGTTACGAACAAAGAAAGCAAGTTTTCCTCA 180
Qy 677 AGATCTTCTACTTC - GGAAGACGAGGCAAGAGCGTGAAGTTTGGGAGGGTAGAGTGTAT 735
Db 181 AGATCTTCTACTCTTTTATGATGCAAGACGTCGAAGTTTTCGAAGGGTAGAGCGAT 240
Qy 736 TTTGGCTGCTCTGMACTGTGGCCCTTGAGTGGGAGGGAAGCTCATATAACACTCGA 795
Db 241 TTTGGCTGCTCTGMACTGTGGCCCTTGAGTGGGAGGGAAGCTCATATAACACTCGA 300
Qy 796 GTTGACTATCGCCCGCTCTGTTGTAACGAGTGCAGTGCATGGGGAATTCGTGGGTGGAG 855
Db 301 GTTGACTATCGCCCGCTCTGTTGTAACGAGTGCAGTGCATGGGGAATTCGTGGGTGGAG 360
Qy 856 GCTGTGCGCACGCTCTCACGTGTGTGGGCCCCCTCACAAAATCGAAACCGCCCGCAGT 915
Db 361 GCTGTGCGCACGCTCTCACGTGTGTGGGCCCCCTCACAAAATCGAAACCGCCCGCAGT 420
Qy 916 CAGAGAGCTTCACTGCCCGTACCATCGAGGGAATGCGTGTATGCTCTTTCACACACCCG 975
Db 421 CAGAGAGCTTCACTGCCCGTACCATCGAGGGAATGCGTGTATGCTCTTTCACACACCCG 480
Qy 976 CTGAATTTTAAGGGAAGTGGCTGCGCAGCCGACTGAACCTCTGGCTGACGG 1027
Db 481 CTGAATTTTAAGGGAAGTGGCTGCGCAGCCGACTGAACCTCTGGCTGACGG 532

RESULT 7
BH197094
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BH197094 495 bp DNA linear GSS 24-OCT-2001
TC3-73F9.TV TC3 Trypanosoma cruzi genomic clone TC3-73F9, genomic survey sequence.
BH197094
BH197094.1 GI:16365250
GSS.
Trypanosoma cruzi
Trypanosoma cruzi
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma; Schizotrypanum.
 1 (bases 1 to 495)
 Kluge, S., Edwards, K.E., Nilsson, D., Bontempi, E.J., Myler, P.,
 Stuart, K., Ghedin, E., El-Sayed, N.M., and Andersson, B.
 Clustering and analysis of BAC-end and GSS sequences from
 Trypanosoma cruzi
 Unpublished (2001)
 Other GSSs: TC3-73F9.TP
 Contact: Bjorn Andersson
 Department of Genetics and Pathology
 Uppsala University
 Rudbeck Laboratory, SE-751 85, Uppsala, Sweden
 Tel: 46 18 471 4107
 Fax: 46 18 471 4808
 Email: bjorn.andersson@genpat.uu.se
 Clones are derived from the Trypanosoma cruzi CL-Brener BAC library
 TC3. For clone availability, please contact Dr. Bjorn Andersson at
 Uppsala University (bjorn.andersson@genpat.uu.se).
 Seq primer: T7
 Class: BAC ends.

FEATURES
 source
 1..495
 /organism="Trypanosoma cruzi"
 /mol_type="genomic DNA"
 /strain="CL Brener"
 /db_xref="taxon:5693"
 /clone="TC3-73F9"
 /clone_lib="TC3"
 /note="Vector: pBelobAC11; Site 1: Hin dIII; Constructed
 for Uppsala University by Marie-Christine Le Paslier in
 the laboratory of Denis Le Paslier at the Centre d'Etude
 du Polymorphisme Humain (CEPH), Paris, France. Briefly,
 Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained
 from Dr. Franco da Silveira) was partially digested with
 Hin dIII. High molecular weight fragments were ligated in
 pBelobAC11 digested with Hin dIII. The average insert
 size is 100 kb. Total clone coverage: approx. 33 X the
 haploid genome."

ORIGIN

Query Match 24.1%; Score 485.4; DB 28; Length 495;
 Best Local Similarity 98.8%; Pred. No. 1.2e-120;
 Matches 489; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 433 GATCGGAGACGCTGGGATATTCGTTGCCGTTGGTGGTACGAGTCCAGTCCGCGC 492
 DB 1 GATCGGAGACGCTGGGATATTCGTTGCCGTTGGTGGTACGAGTCCAGTCCGCGC 60

QY 493 GCGAAGATACTCGAGTATCAATGGGGAGCCCGTGTCACTGAAGCAATTTTCCG 552
 DB 61 GCGAAGATACTCGAGTATCAATGGGGAGCCCGTGTCACTGAAGCAATTTTCCG 120

QY 553 GCGAATGAAGGAATGCACAAATTCCTTTGGCGGTGCAGGTGTGCCATTGTG 612
 DB 121 GCGAATGAAGGAATGCACAAATTCCTTTGGCGGTGCAGGTGTGCCATTGTG 180

QY 613 GCGTCCAGCGGAATCTGTGTACCTGTGCAGTTACGAACAAAGAGCAAGTTT 672
 DB 181 GCGTCCAGCGGAATCTGTGTACCTGTGCAGTTACGAACAAAGAGCAAGTTT 240

QY 673 TCCAAGATCTTCTACTCGGAAGACGAGGCAAGAGCTGGAAGTGTGGGAGGTAGGAGT 732
 DB 241 TCTAAGATCTTCTACTCGGAAGATGATGCAAGAGCTGGAAGTGTGGGAGGTAGGAGC 300

QY 733 GATTTTGGCTGCTGTAACCTGTGGCCCTTGTAGTGGAGGGGAAGCTCATATAACACT 792
 DB 301 GATTTTGGCTGCTGTAACCTGTGGCCCTTGTAGTGGAGGGGAAGCTCATATAACACT 360

QY 793 CGAGTTGACTATCGCGCGCTCTGTGTACAGTGCAGTGCATGGGGAATTCGTGGTG 852
 DB 361 CGAGTTGACTATCGCGCGCTCTGTGTACAGTGCAGTGCATGGGGAATTCGTGGTG 420

QY 853 GAGGCTGTGCGGCAACGCTCTACGCTGTGTGGGGCCCTCACCAAAATCGAACCCCGC 912

Db 421 GAGGCTGTGCGGACGCTCTCAGCTGTGTGGGGCCCTCACCAAAATCGAACCCCGC 480
 QY 913 AGTCAGACGAGCTTC 927
 Db 481 AGTCAGACGAGCTTC 495

RESULT 8
 BH844627
 LOCUS
 DEFINITION
 TC3-50N19.TV TC3 Trypanosoma cruzi genomic clone TC3-50N19, genomic
 survey sequence.
 BH844627
 BH844627.1 GI:21414345
 GSS.
 Trypanosoma cruzi
 Trypanosoma cruzi
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma; Schizotrypanum.
 1 (bases 1 to 526)
 Myler, P.J., Aggarwal, G., Fazelinia, G., Mack, J., Marty, A.,
 Munden, H., Nelson, S., Pentony, M., Rinta, J., Robertson, L.,
 Seyler, A., Siek, E., Stuart, K., Vogt, C., Worthey, E., El-Sayed, N.M.,
 Ghedin, E. and Andersson, B.
 Trypanosoma cruzi CL-Brener TC3 BAC-end sequencing
 Unpublished (2001)
 Other GSSs: TC3-50N19.TP TC3-50N19.TP.1
 Contact: Peter Myler
 Seattle Biomedical Research Institute
 4 Nickerson Street, Seattle, WA 98109, USA
 Tel: 206 284 8846
 Fax: 206 284 0313
 Email: mylerp@bri.org
 Clones are derived from the Trypanosoma cruzi CL-Brener BAC library
 TC3. For clone availability, please contact Dr. Bjorn Andersson at
 Uppsala University (bjorn.andersson@genpat.uu.se).
 Seq primer: T7
 Class: BAC ends.

FEATURES
 Location/Qualifiers
 1..526
 /organism="Trypanosoma cruzi"
 /mol_type="genomic DNA"
 /strain="CL Brener"
 /db_xref="taxon:5693"
 /clone="TC3-50N19"
 /clone_lib="TC3"
 /note="Vector: pBelobAC11; Site 1: Hin dIII; Constructed
 for Uppsala University by Marie-Christine Le Paslier in
 the laboratory of Denis Le Paslier at the Centre d'Etude
 du Polymorphisme Humain (CEPH), Paris, France. Briefly,
 Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained
 from Dr. Franco da Silveira) was partially digested with
 Hin dIII. High molecular weight fragments were ligated in
 pBelobAC11 digested with Hin dIII. The average insert
 size is 100 kb. Total clone coverage: approx. 33 X the
 haploid genome."

ORIGIN

Query Match 22.9%; Score 460; DB 28; Length 526;
 Best Local Similarity 92.4%; Pred. No. 1e-113;
 Matches 484; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 374 AGCTTTACGTCCTCGTTGGAAGCTCAACAGTTCCGAGGAGCTACTGGACGTGCGATGGTG 433
 DB 1 AGCTTTACGTCCTCGTTGGAAGCTCAACAGTTCCGAGGAGCTACTGGACGTGCGACCTG 60

QY 434 ATCGGAGAGACTGGGATATTCGTTGCCGTTGGTGGAGGTACGAAGTCCACTCGCGGGCG 493
 DB 61 ATGGAAGCAGACTGGGACATTCGTTGCCGTTGGTGGAGGTACGAAGTCCACTCGCGGGCG 120

QY 494 GCAAGATACTCGAGTATCAATGGGGAGCCCGCTGTCACTGAAGCAATTTTCCCGG 553

FEATURES
source

Location/Qualifiers
1. .467
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/strain="CL Brener"
/db_xref="taxon:5693"
/clone="TC3-51M23"
/clone_lib="TC3"
/note="Vector: pBelOBAC11; Site 1: Hin dIII; Constructed for Uppsala University by Marie-Christine Le Paslier in the laboratory of Denis Le Paslier at the Centre d'Etude du Polymorphisme Humain (CEPH), Paris, France. Briefly, Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained from Dr. Franco da Silveira) was partially digested with Hin dIII. High molecular weight fragments were ligated in pBelOBAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 33 X the haploid genome."

ORIGIN

Query Match 18.5%; Score 378.2; DB 28; Length 467;
Best Local Similarity 89.5%; Pred. No. 1.8e-91;
Matches 418; Conservative 0; Mismatches 48; Indels 1; Gaps 1;
QY 557 AAATCGAAGGATGACACAAATCAATTTCTTGGCGGTGCGAGGTGTGCCATTGTGGCGT 616
Db 2 ACATGGAAGGAATACCCCAAGCAATTTCTT-TTTTTCAGGTGTGGCCACTGTGGCGT 60
QY 617 CCAACGGGAATCTGTGTACCTGTGACAGTTTACGACAAAGAAAGCAAGTATTTTCCA 676
Db 61 CCAACGGGAATCTGTGTACCTGTGACAGTTTACGACAAAGAAAGCAAGTATTTTCCA 120
QY 677 AGATCTTCTACTCGGAAGAAGAGCGGCAAGAGCTGGAAGTTTGGGGAGGGTAGGAGTGATT 736
Db 121 AGATCTTCTTATTCGGCAGATGAGGCAAGAGCTGGAATTTCTCAGAGGGTAGGAGGATT 180
QY 737 TTGGCTGTCTGAACCTGTGCCCTTCAGTGGGAGGGAGAGCTCATCAAAACACTCGAG 796
Db 181 TTGGCTGTCTGAACCTGTGGTCTTGAGTGGGAAATAAATTCATCAAAACACTCGAG 240
QY 797 TTGACTATCGCCGCGCTGTGGTGTACAGTCCAGTGCATGCGGGAATTCGTGGTGGAGG 856
Db 241 TTGACTATCGCCGCGCTGTGGTGTACAAATTCGGTGTACATGGAACCGTGGTGGAGG 300
QY 857 CTGTGGCAGCTCTCAGTGTGTGGGCCCCCTCACAAAATCGAACCCCGGCGAGTC 916
Db 301 CTGTGGCAGCTCTCAGTGTGTGGGCCCCCTCACAAAATCGAACCCCGGCGAGTC 360
QY 917 AGAGCAGCTTCACTGCGCGTCAACCATCGAGGAATGCTGTATGCTCTTCACACACCCGC 976
Db 361 AGAGCAGCTTCACTGCGCGTCAACCATCGAGGAATGCTGTATGCTCTTCACACACCCGC 420
QY 977 TGAATTTAAGGAAGTGGCTGCGCAGCGAGTGAACCTCTGGCTG 1023
Db 421 TGAATTTAAGGAAGTGGCTGCGCAGCGAGTGAACCTCTGGCTG 467

RESULT 11

AQ911191 412 bp DNA linear GSS 09-JAN-2001
LOCUS GSSTC03225 Trypanosoma cruzi random genomic library Trypanosoma
DEFINITION cruzi genomic clone G40M7, genomic survey sequence.
ACCESSION AQ911191
VERSION AQ911191.3 GI:10136322
KEYWORDS GSS.

SOURCE
ORGANISM

Trypanosoma cruzi
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
1 (bases 1 to 412)

REFERENCE

AQUERO, F., Verdun, R., Frasch, A. C. C. and Sanchez, D. O.
AUTHORS A random sequencing approach for the analysis of the trypanosoma
TITLE cruzi genome: general structure, large gene and repetitive DNA

JOURNAL
MEDLINE
PUBMED
COMMENT

families, and gene discovery
Genome Res. 10 (12), 1996-2005 (2000)
20568489
11116094
On Sep 14, 2000 this sequence version replaced gi:9370791.
Contact: Sanchez D.O.
Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral
San Martin)
Av. Gral Paz S/N, INTI, Edificio 24, B 1650 KNA, San Martin, Buenos
Aires, Argentina
Tel: (54-11) 4580/7255/7
Fax: (54-11) 4752-9639
Email: dsanchez@fib.unsam.edu.ar
Sequences were basecalled with phred and vector was masked with
crossmatch (see http://genome.washington.edu). Sequences were then
trimmed from both ends to remove low quality bases and masked
vector.
Seq primer: T7
Class: shotgun.

FEATURES
source

Location/Qualifiers
1. .412
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/strain="CL-Brener"
/db_xref="taxon:5693"
/clone="G40M7"
/cell_type="epimastigote"
/clone_lib="Trypanosoma cruzi random genomic library"
/note="Vector: pBS(-) (Stratagene); T. cruzi DNA was
randomly sheared using a nebulizer and the 1 to 2 Kb range
was gel purified and cloned into the dephosphorylated
HincII site of the vector."

ORIGIN

Query Match 18.5%; Score 372; DB 28; Length 412;
Best Local Similarity 93.9%; Pred. No. 8.2e-90;
Matches 387; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
QY 689 CGGAAGCAGGGGCAAGCCTGGAAGTTTGGGAGGGTAGGAGTATTTGGCTGCTCG 748
Db 1 CGGAAGCAGGGGCAAGCCTGGAAGTTTGGGAGGGTAGGAGGCTTTGGCTGCTCGC 60
QY 749 AACCTGTGGCCCTTGAAGTGGGAGGGAGGAGCTCATATAAACACTCGAGTTGACTATCGCC 808
Db 61 AACCTGTGGCCCTTGAAGTGGGAGGGAGGAGCTCATATAAACACTCGAGTTGACTGGAAC 120
QY 809 GCCCTCTGGTACGAGTCCAGTGCATGGGGAATTGCTGGGTGGAGGCTGTGCGCACGC 868
Db 121 GCCCCTCGGTGTACGAGTCCAGTGCATGGGGAATACGTGGGTGGAGGCTGTGCGCACGC 180
QY 869 TCTCACGCTGTGGGGCCCTCACAAAATCGAACCCCGGAGCTCAGAGCAGCTTCA 928
Db 181 TCTCGCTGTGGGGCCCTCACAGCATCGAATATGCCCGGAGCTCAGAGCAGCTTCA 240
QY 929 CTGCGGTGACCATCGAGGGGAATCGGTGTATGCTCTTTCACACCCCGCTGAAATTTAAAG 988
Db 241 CTGCGGTGACCATCGAAGGAATCGGTGTATGCTCTTTCACACCCCGCTGAAATTTAAAG 300
QY 989 GAAGGTGGCTGCCGACCCGACCTCTGGCTGAGGATACCAAGGATACCAAGGATTTATAAG 1048
Db 301 GATGTGGCTGCCGACCCGACCTCTGGCTGAGGATACCAAGGATTTATAAG 360
QY 1049 TTGGGCAAGTATCATTGGTGTGATGAAAATTCGCGCTTACAGCTCCGCTCTGTA 1100
Db 361 TTGGGCAAGTATCATTGGTGTGATGAAAATTCGCGCTTACAGCTCCGCTCTGTA 412

RESULT 12

BH191346
LOCUS
DEFINITION
ACCESSION
BH191346
TC3-40G16.TF TC3 Trypanosoma cruzi genomic clone TC3-40G16, genomic
survey sequence.
AUTHORS
TITLE

VERSION BH191346.1 GI:16358622
KEYWORDS
SOURCE Trypanosoma cruzi
ORGANISM Trypanosoma cruzi
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
1 (bases 1 to 471)
REFERENCE
AUTHORS Ghedin, E., Malek, J., Shetty, J., Khalak, H., Koo, H., Myler, P.,
Stuart, K., Andersson, B. and El-Sayed, N.M.
TITLE Use of BAC end sequences from Trypanosoma cruzi CL-Brener TC3
library for gene discovery and map construction
JOURNAL Unpublished (2001)
COMMENT Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma cruzi CL-Brener BAC library
TC3. For clone availability, please contact Dr. Bjorn Andersson at
Uppsala University (bjorn.andersson@genpat.uu.se).
Seq primer: M13 For
Class: BAC ends.
Location/Qualifiers
1. .471
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/strain="CL Brener"
/db_xref="taxon:5693"
/clone="TC3-40G16"
/clone_lib="TC3"
/note="Vector: pBelOBAC11; Site 1: Hin dIII; Constructed
for Uppsala University by Marie-Christine Le Paslier in
the laboratory of Denis Le Paslier at the Centre d'Etude
du Polymorphisme Humain (CEPH), Paris, France. Briefly,
Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained
from Dr. Franco da Silveira) was partially digested with
Hin dIII. High molecular weight fragments were ligated in
pBelOBAC11 digested with Hin dIII. The average insert
size is 100 kb. Total clone coverage: approx. 33 X the
haploid genome."

Query Match 18.0%; Score 360.8; DB 28; Length 471;
Best Local Similarity 85.7%; Pred. No. 1e-86;
Matches 401; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 374 AGCTTTACGTCCTCGGTGGAAGCTACACAGTTTCGAGGAGCTACTGGACGTGCGCATGGTG 433
DB 1 AGCTTTACGTCCTCGGTGGAAGCTACCAATAAATCGAAAAAATCTACTGGACTTGGCAGCCTG 60

QY 434 ATGCGAGAGACTGGGATATTCTGTTGCCGTTGGTGAGTTCAGGATCAGGATCCTCGGGCG 493
DB 61 ATGGAAGCGACTGGGATATTCTGTTGCCGTTGGTGAGGTTACGAATGCTCACTGCAGAGC 120

QY 494 GCAAGATACTCGAGTATCAATGCGGGAGCCCGTGTCACTGAAGGAATTTTCCCGG 553
DB 121 GCAAGACAATCGGAATATCAATGCGGGAGTCCCGTGTCACTGAAGGAATTTTCCCGG 180

QY 554 CGGAATGGAAGGATGACACAAATTTCTTGGCGGTGCGAGTGTGCAATTTGTG 613
DB 181 CATACATGGAAGATACCTACAAAGCAATTTCTCGCGGTGCGAGTGTGCACTGTG 240

QY 614 CGTCCAAACGGGAATCTTGTGTACCCCTGTGAGGTTACGAACAAAAGACGAAGTTTTTT 673
DB 241 CGTCCAAATGGGAATCTTGTGTACCCCTGTGAGGTTACGGACATGAAAAAGCAAGTTTTTT 300

QY 674 CCAAGATCTTCTACTCGGAAGACGAGGCAAGCGTGGGAAGTTTGGGGAGGGTAGGAGTG 733
DB 301 CCAAGATCTTTTATTCGGCAGATGAGGGCAAGACGTGGAAATTTCTCAGAGGGTAGGAGC 360

QY 734 ATTTTGGCTGCTCTGAACCTGTGGCCCTTTGAGTGGGAGGGGAAGCTCATATAAACACTC 793

DB 361 ATTTTGGCTGCTCCGAACCTGTGGTCTTGAAGTGGAAAAATAAATTCATCGTAAACACC 420
QY 794 GAGTTGACTATCCGCCCGCTCTGGTGTACAGTCCAGTGCACATGGGGA 841
DB 421 GAGTTGACCGGGCGGCCCTGTGGTGTACAAATTCGGTGACATGGAGA 468

RESULT 13
LOCUS AQ910653
DEFINITION GSSTC03489 Trypanosoma cruzi random genomic library Trypanosoma
cruzi genomic clone G43G9, genomic survey sequence.
ACCESSION AQ910653
VERSION AQ910653.3 GI:10137022
KEYWORDS GSS.
SOURCE Trypanosoma cruzi
ORGANISM Trypanosoma cruzi
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
1 (bases 1 to 366)
REFERENCE
AUTHORS Aguero, F., Verdun, R., Frasch, A.C.C. and Sanchez, D.O.
TITLE A random sequencing approach for the analysis of the trypanosoma
cruzi genome: general structure, large gene and repetitive DNA
families, and gene discovery
JOURNAL Genome Res. 10 (12), 1996-2005 (2000)
MEDLINE 20568489
PUBMED 11116094
COMMENT On Sep 14, 2000 this sequence version replaced gi:9376713.
Contact: Sanchez D.O.
Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral
San Martin)
Av. Gral Paz S/N, INTI, Edificio 24, B 1650 KNA, San Martin, Buenos
Aires, Argentina
Tel: (54-11) 4580/7255/7
Fax: (54-11) 4752-9639
Email: dsanchez@lib.unsam.edu.ar
Sequences were basecalled with phred and vector was masked with
crossmatch (see http://genome.washington.edu). Sequences were then
trimmed from both ends to remove low quality bases and masked
vector.
Seq primer: T7
Class: shotgun.
Location/Qualifiers
1. .366
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/strain="CL-Brener"
/db_xref="taxon:5693"
/clone="G43G9"
/cell_type="epimastigote"
/clone_lib="Trypanosoma cruzi random genomic library"
/note="Vector: pBS(-) (Stratagene); T. cruzi DNA was
randomly sheared using a nebulizer and the 1 to 2 Kb range
was gel purified and cloned into the dephosphoryated
HincII site of the vector"

ORIGIN
Query Match 17.0%; Score 342; DB 28; Length 366;
Best Local Similarity 95.9%; Pred. No. 1.1e-81;
Matches 351; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1036 CGCATTTATACGTTGGGCAAGTATCCATTGGTGATGAAAAATTCGCCCTACAGCTCCGTC 1095
DB 1 CGCATTTATACGTTGGGCAAGTATTCATTGGTGATGAAAAATTCGCCCTACAGCTCCGTC 60

QY 1096 CTGTACAAGGATGATAAGCTGTACTGTTTGCATGAGATCAACAGTAACGAGGTGTACAGC 1155
DB 61 CTGTACAAGGATGATAAGCTGTACTGTTTGCATGAGATCAACAGTAACGAGGTGTACAGC 120

QY 1156 CTGTGTTTTCGGCGCTGTTGGCGAGCTACCGATCATTAATCATGCTGCGAGTCTCTGG 1215
DB 121 CTGTGTTTTCGGCGCTGTTGGCGAGCTACCGATCATTAATCATGCTGCGAGTCTCTGG 180

QY 1216 AAGATTGGACAGCCACTGTCCAGCATTTGACCCCTGCTGATCCAGCCGCTTGGTCG 1275
 Db 181 AAGATTGGACAGCCACTGTCCAGCATTTGACCCCTGCTGATCCAGCCACTTGGTCG 240
 QY 1276 TCAGAGCGTGTGTGGTCCCGCTGTCACACCGTGTGGTCTTGTGGCTTTTGTGCGAC 1335
 Db 241 TCAGAGCGTGTGTGGTCCCGCTGTCACACTACGGTGTGCTGTGCTGTGTGTGCGAC 300
 QY 1336 AGTGCCACCAACCAAGTGGAGGATGCTACCGTGGTCAACGACGACGCAAAAT 1395
 Db 301 AATGCCACCAACCAAGTGGAGGATGCTACCGTGGTCAACGACGACGCAAAAT 360
 QY 1396 GCGGAG 1401
 Db 361 GCGGAG 366

RESULT 14
 BH844382
 LOCUS
 DEFINITION TC3-53G10.TP TC3 Trypanosoma cruzi linear GSS 13-JUN-2002
 survey sequence.
 ACCESSION BH844382
 VERSION
 KEYWORDS BH844382.1 GI:21413850
 SOURCE GSS.
 ORGANISM Trypanosoma cruzi

REFERENCE
 AUTHORS Trypanosoma cruzi
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma; Schizotrypanum.
 1 (bases 1 to 367)

Myler,P.J., Aggarwal,G., Fazelinia,G., Mack,J., Marty,A.,
 Munden,H., Nelson,S., Pentony,M., Rinta,J., Robertson,L.,
 Seyler,A., Sisk,E., Stuart,K., Vogt,C., Worthey,E., El-Sayed,N.M.,
 Ghedin,E. and Anderson,B.
 Trypanosoma cruzi CL-Brener TC3 BAC-end sequencing
 Unpublished (2001)

Other_GSSs: TC3-53G10-TV
 Contact: Peter Myler
 Seattle Biomedical Research Institute
 4 Nickerson Street, Seattle, WA 98109, USA
 Tel: 206 284 8846
 Fax: 206 284 0313
 Email: mylerpj@bri.org
 Clones are derived from the Trypanosoma cruzi CL-Brener BAC library
 TC3. For clone availability, please contact Dr. Bjorn Andersson at
 Uppsala University (bjorn.andersson@genpat.uu.se).
 Seq primer: SP6
 Class: BAC ends.

FEATURES
 Location/Qualifiers
 1..367
 /organism="Trypanosoma cruzi"
 /mol_type="genomic DNA"
 /strain="CL Brener"
 /db_xref="taxon:5693"
 /clone="TC3-53G10"
 /clone_lib="TC3"
 /note="Vector: pBeloBAC11; Site 1: Hin dIII; Constructed
 for Uppsala University by Marie-Christine Le Paslier in
 the Laboratory of Denis Le Paslier at the Centre d'Etude
 du Polymorphisme Humain (CEPH), Paris, France. Briefly,
 Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained
 from Dr. Franco da Silveira) was partially digested with
 Hin dIII. High molecular weight fragments were ligated in
 pBeloBAC11 digested with Hin dIII. The average insert
 size is 100 kb. Total clone coverage: approx. 33 X the
 haploid genome."

ORIGIN

Query Match 16.4%; Score 330.6; DB 28; Length 367;
 Best Local Similarity 95.6%; Pred. NO. 1.5e-78;
 Matches 351; Conservative 0; Mismatches 14; Indels 2; Gaps 1;

QY 407 CCGAGGAGCTACTGCACTGCGATGTTGATGCGAGAGACTGGGATATTTCTGCTTGCCTTG 466
 Db 1 CGAGGAGCTACTGCACTGCGATGTTGATGCGAGAGACTGGGATATTTCTGCTTGCCTTG 60
 QY 467 GTGAGGTACGAAGTCCACTGCGGGCGGCAAGATAAATCGCGAGTATCAAAATGGGGAGCC 526
 Db 61 GTGAGGTACGAAGTCCACTGCGGGCGGCAAGATAAATCGCGAGTATCAAAATGGGGAGCC 120
 QY 527 CCGTGTCACTGAAGGAATTTTCCCGGGCGGAATGGAAGGATGCAACAAATCAATTTTC 586
 Db 121 CCGTGTCACTGAAGGAATTTTCCCGGGCGGAATGGAAGGATGCAACAAATCAATTTTC 180
 QY 587 TTGGCG--GTGCGAGTGTTCCTTGTGGCGTCCAAACGGGAATCTTGTGTACCTGTGCA 644
 Db 181 TTGGCGTTTCTTGTGTGGCGTCCAAACGGGAATCTTGTGTACCTGTGCA 240
 QY 645 GGTTCGACAAACAAAGCAAGTTTTTCCAGATCTTCTACTCGGAAGACGAGGCAA 704
 Db 241 GGTTCGACAAACAAAGCAAGTTTTTCCAGATCTTCTACTCGGAAGATGATGGCAA 300
 QY 705 GACGTGGAAGTTTGGGGAGGTAGGAGTGAATTTTGGCTGCTCTGAACCTGTGGCCCTTGA 764
 Db 301 GACGTGGAAGTTTGGGGAGGTAGGAGTGAATTTTGGCTGCTCTGAACCTGTGGCCCTTGA 360
 QY 765 GTGGGAG 771
 Db 361 GTGGGAG 367

RESULT 15
 A0911046
 LOCUS
 DEFINITION GSSr02480 Trypanosoma cruzi random genomic library Trypanosoma
 A0911046
 ACCESSION
 VERSION
 KEYWORDS A0911046.3 GI:10133494
 SOURCE GSS.
 ORGANISM Trypanosoma cruzi

Trypanosoma cruzi
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma; Schizotrypanum.
 1 (bases 1 to 364)

Aguero,F., Verdun,R., Fraach,A.C.C. and Sanchez,D.O.
 A random sequencing approach for the analysis of the trypanosoma
 cruzi genome: general structure, large gene and repetitive DNA
 families, and gene discovery
 Genome Res. 10 (12), 1996-2005 (2000)
 20568489
 11116094

On Sep 14, 2000 this sequence version replaced gi:9371848.
 Contact: Sanchez D.O.

Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral
 San Martin)
 Av. Gral Paz S/N, INTI, Edificio 24, B 1650 KNA, San Martin, Buenos
 Aires, Argentina
 Tel: (54-11) 4580/7255/7
 Fax: (54-11) 4752-9639
 Email: dsanchez@iib.unsam.edu.ar

Sequences were basecalled with phred and vector was masked with
 crossmatch (see http://genome.washington.edu). Sequences were then
 trimmed from both ends to remove low quality bases and masked
 vector.

Seq primer: T7
 Class: shotgun.

FEATURES
 source

1..364
 /organism="Trypanosoma cruzi"
 /mol_type="genomic DNA"
 /strain="CL-Brener"
 /db_xref="taxon:5693"
 /clone="G28D17"
 /cell_type="epimastigote"
 /clone_lib="Trypanosoma cruzi random genomic library"

/notes="Vector: pBS(-) (Stratagene); T. cruzi DNA was randomly sheared using a nebulizer and the 1 to 2 Kb range was gel purified and cloned into the dephosphorylated HincII site of the vector"

ORIGIN		
Query Match		15.7%; Score 315.2; DB 28; Length 364;
Best Local Similarity		93.4%; Pred. No. 2.3e-74;
Matches 340; Conservative		0; Mismatches 23; Indels 1; Gaps 1;
Qy	942	CGAGGGAATGCGTTATGCTCTTCACACACCGCTGA-ATTTTAAAGGAAGTGGCTGC 1000
Db	1	CGAAGGAATGCGTGTGATGCTCTTCACACACCGCTGTATTTTAAAGGAATGTTGGCTGC 60
Qy	1001	GCACCGACTGAACCTCTGGCTGACGGATAACACGCGCANTTTATAACGTTGGCAAGTAT 1060
Db	61	GCACCGACTGCAGCTTTGGCTGACGCAACACGCGCATTTATAACGTTGGCAAGTAT 120
Qy	1061	CCATTGGTGATGAAAAATTCCGCCCTACAGCTCCGTCCTGTACAAGGATGATAAGCTGTACT 1120
Db	121	CCATTGGTGATGAAAAATTCCGCCCTACAGCTCCGTCCTGTACAAGGATGATAAGCTGTACT 180
Qy	1121	GTTTGCATGAGATCAACAGTAAACAGGTGTACAGCCCTGTTTTCGCGCCTGGTTGGCG 1180
Db	181	GTTTGCATGAGATCAACAGTAAACAGGTGTACAGCCCTGTTTTCGCGCCTGGTTGGCG 240
Qy	1181	AGCTACGGATCAATAATCAGTGTCTGCAGTCCCTGGGAAGAAATGGGACAGCCACTGTCCA 1240
Db	241	AGCTACGGCTCATCAATCCGTGGTGGTTCCTGGGAAGTGGGACAGCCACTGTCCA 300
Qy	1241	GCATTTGCACCCCTGCTGATCCAGCCGCTTCGTCGTGCAGAGCGTGTGTGTGTCCTCCGCTG 1300
Db	301	GCAATTGCACCCCTGCTGATTCGTGCTCTTCGTCTGTCGTCGTCGAGCGTGTGTGTGTCCTCCGCTG 360
Qy	1301	TCAC 1304
Db	361	TCAC 364

Search completed: August 9, 2004, 00:42:38
Job time : 7678 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 09:32:50 ; Search time 14 seconds
(without alignments)

2488.210 Million cell updates/sec

Title: US-10-086-913-2

Perfect score: 3517

Sequence: 1 MGSSHHHHHSSGLVPRGSH.....EAMGSSSGSRSTPGSGC 669

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2681	76.2	1162	1 TCNA TRYCR	P23253 trypanosoma
2	832	23.7	714	1 GP85 TRYCR	Q03877 trypanosoma
3	815	23.2	752	1 8511 TRYCR	P18269 trypanosoma
4	394.5	11.2	404	1 NANH CLOSO	P15698 clostridium
5	388	11.0	382	1 NANH CLOPE	P10481 clostridium
6	344	9.8	381	1 NANH SALTY	P29768 salmonella
7	291.5	8.3	175	1 8513 TRYCR	P18271 trypanosoma
8	226.5	6.4	1035	1 NANA STRPN	Q59959 streptococc
9	208	5.9	240	1 8512 TRYCR	P18270 trypanosoma
10	195	5.5	1014	1 NANH CLOSE	P29767 clostridium
11	191.5	5.4	266	1 NANH BACFR	P31206 bacteroides
12	171.5	4.9	647	1 NANH MICVI	Q02834 micromonosp
13	157	4.5	697	1 NANE STRPN	Q54727 streptococc
14	148.5	4.2	1861	1 APU THETU	P38536 t amylopull
15	142.5	4.1	409	1 NERI MOUSE	Q35657 mus musculus
16	139.5	4.0	409	1 NERI RAT	Q99pw3 rattus norv
17	134	3.8	781	1 NANH VIRCH	P37060 vibrio chol
18	129	3.7	415	1 NERI HUMAN	Q99519 homo sapien
19	125	3.6	1222	1 SOR3 HUMAN	Q9upw3 homo sapien
20	124.5	3.5	380	1 NERI HUMAN	Q9y3r4 homo sapien
21	122.5	3.5	3178	1 YS89 CAEEL	Q09624 caenorhabdi
22	121.5	3.5	611	1 WDRI CAEEL	Q11176 caenorhabdi
23	121	3.4	962	1 GUNA PSEFL	P10476 pseudomonas
24	119	3.4	379	1 NER2 MOUSE	Q9jnh3 mus musculus
25	118.5	3.4	1277	1 CAML FUGRU	Q98902 fugu rubrip
26	117.5	3.3	95	1 DAFT TRYCR	Q26327 trypanosoma
27	117.5	3.3	1848	1 CBPA CLOCL	P38058 clostridium
28	116.5	3.3	1219	1 SOR3 MOUSE	Q8v151 mus musculus
29	116	3.3	853	1 PHS1 DICDI	Q00766 dictyosteli
30	114	3.2	1159	1 SOR2 MOUSE	Q09pr5 mus musculus
31	113.5	3.2	618	1 MUTL SALTI	Q8z187 salmonella
32	112.5	3.2	618	1 MUTL SALTY	P14161 salmonella
33	112.5	3.2	693	1 LYS4 YEAST	P49367 saccharomyc

ALIGNMENTS				
RESULT 1				
TCNA_TRYCR	ID	TCNA_TRYCR	STANDARD;	PRT; 1162 AA.
DT	P23253;	AC	01-NOV-1991 (Rel. 20, Created)	
DT	01-NOV-1991 (Rel. 20, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Sialidase (EC 3.2.1.18) (Neuraminidase) (NA) (Major surface antigen) .			
GN	TCNA.			
OS	Trypanosoma cruzi.			
OC	Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.			
OX	NCBI_TaxID:5693;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Silvio X-10/4;			
RX	MEDLINE=91277609; PubMed=1711561;			
RA	Pereira M.E.A., Mejia J.S., Ortega-Barria E., Matzilevich D.,			
RA	Prioli R.P.;			
RT	"The Trypanosoma cruzi neuraminidase contains sequences similar to			
RT	bacterial neuraminidases, YWTD repeats of the low density lipoprotein			
RT	receptor, and type III modules of fibronectin.";			
RL	J. Exp. Med. 174:179-191(1991).			
RN	[2]			
RP	SUBCELLULAR LOCATION.			
RA	MEDLINE=91376547; PubMed=1896773;			
RA	Prioli R.P., Mejia J.S., Aji T., Alkawa M., Pereira M.E.A.;			
RA	"Trypanosoma cruzi: localization of neuraminidase on the surface of			
RL	trypanomastigotes.";			
RL	Trop. Med. Parasitol. 42:146-150(1991).			
CC	-I- FUNCTION: Developmentally regulated neuraminidase implicated in			
CC	parasite invasion of cells.			
CC	-I- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-, alpha-(2->8)-glycosidic linkages of terminal sialic residues in			
CC	oligosaccharides, glycoproteins, glycolipids, colominic acid and			
CC	synthetic substrates.			
CC	-I- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor			
CC	(potential).			
CC	-I- DEVELOPMENTAL STAGE: Maximal activity in trypomastigotes, minimum			
CC	in epimastigotes and not detectable in amastigotes.			
CC	-I- MISCELLANEOUS: The variable lengths of the long tandem repeat			
CC	domain could account in part for the polymorphism of the TCNA			
CC	protein.			
CC	-I- SIMILARITY: Belongs to family 33 of glycosyl hydrolases.			
CC	-I- SIMILARITY: Contains 3 BNR repeats.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M61732; AAA30255.1; --			
DR	PIR; JH0557; JH0557.			
DR	InterPro; IPR002860; GH_BNR.			

```
DR InterPro; IPR008377; Sialidase_trypan.
DR Pfam; PF02012; BNR_2.
DR PRINTS; PR01803; TCSIALIDASE.
KW Hydrolase; Glycosidase; Glycoprotein; Repeat; GPI-anchor;
KW Phosphorylation.
FT DOMAIN 1 457 CYS-RICH.
FT REPEAT 23 34 BNR 1.
FT REPEAT 163 174 BNR 2.
FT REPEAT 209 220 BNR 3.
FT DOMAIN 458 598 FIBRONECTIN TYPE-III.
FT DOMAIN 589 1120 44 X 12 AA TANDEM REPEATS, LTR DOMAIN.
FT CARBOHYD 342 342 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1125 1125 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1162 AA; 120032 MW; 07049221897C6A40 CRC64;

Query Match
Best Local Similarity 76.2%; Score 2681; DB 1; Length 1162;
Matches 525; Conservative 22; Mismatches 50; Indels 10; Gaps 4;

QY 66 MVAIADARYETSDNSLIDTVAKSVDDGETWETQIAIKNSRASSVSRVVDPTVIVKGNK 125
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MVAIADARYETSSNSLIDTVAKSVDDGETWETQIAIKNSRASSVSRVVDPTVIVKGNK 60
QY 126 LYVLGVSNSRSYWTSHGDRDWDILLAVGEVTKTAGGKITASIKWSPVSLKEFFPA 185
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 LYVLGVSYSRSYSSWSSHGDRDWDILLAVGEVTKTAGGKITASIKWSPVSLKEFFPA 120
QY 186 EMEGHNTQFLLGGAGVAIVASGNLVVPQVQVTKKKQVFSKIFSEDEGKTWFGEGRSD 245
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 EMEGHNTQFLLGGAGVAIVASGNLVVPQVQVTKKKQVFSKIFSEDDGKTWFGEGRSD 180
QY 246 FGCSEPVALEWEGKLIINTRVDYRRRLVYBSSDMGNSWEAVGTLRSVMGSPSKNQPGS 305
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 FGCSEPVALEWEGKLIINTRVDYRRRLVYBSSDMGNSWEAVGTLRSVMGSPSKNQPGS 240
QY 306 QSSFTAVTIEGRVWMLFTHPLNFKRWLRDLNLWLTNDORIYNVQVSGISDENSAYSSV 365
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241 QTSFTAVTIEGRVWMLFTHPLNFKRCVRDLNLWLTNDORIYNVQVSGISDENSAYSSV 300
QY 366 LYKDDKLYCLHEINSNEVSLFARLVGELRIKSVLOSKNWDSHLSSTCTPADPAASS 425
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 LYKDDKLYCLHEINTDEVSLFARLVGELRIKSVLSRWKMW----TATCPAFAPLLIQ 356
QY 426 SERCCGPAVTVG-----LVGFLSHSATKTEWEDAYRCVNASTANAERVPNGLKFAVGCGG 481
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
357 PLARQRVVVVVPLSPRLVLLAFRCRQLPK-RMGGSYRCVNASTANAERVRNGLKFAVGCGG 415
QY 482 ALWPVSQOQGNQRYHFANHAFTLVASVTIHEVPSVASPLLGASLDSSGGKLLGLSYDEK 541
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
416 ALWPVSQOQGNQRYHFANHAFTLVASVTIHEAPPAASPLLGASLDSSGGKLLGLSYDEK 475
QY 542 HQWQPIYGSTPTPTGSEWNGKRYHYVLTMANKIGSVYIDGFBLEGGQTVPVPGRTPTDI 601
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
476 HQWQPIYGSTPTPTGSEWTKRYHLVLTWANKIGSVYIDGELLEGGQTVPVPGRTPTDI 535
QY 602 SHEYVGGYGRSDMPTTSHVTVNNVLLYN-ROLNABEIRTLFLSODLIGTEAHMGSSSGSS 660
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
536 SHEYVGGYGRSDMPTTSHVTVNNVLLYNRRQLNTEIRTLFLSODLIGTEAHMDSDDSS 595
QY 661 ERSTPGS 667
Db |||||:
596 AHSTPGT 602

RESULT 2
GP85_TRYCR
ID GP85_TRYCR STANDARD; PRT; 714 AA.
AC Q03877;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 85 kDa surface antigen precursor.
```

```
GN GP85.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RX MEDLINE=92107221; PubMed=1762630;
RA Takle G.B., Cross G.A.M.;
RT "An 85-kilodalton surface antigen gene family of Trypanosoma cruzi encodes polypeptides homologous to bacterial neuraminidases.";
RL Mol. Biochem. Parasitol. 37:57-64(1989).
RL [2].
RP SEQUENCE OF 42-461 FROM N.A.
RC STRAIN=Y;
RX MEDLINE=90136716; PubMed=2693963;
RA Takle G.B., Young A., Snary D., Hudson L., Nicholls S.C.;
RT "Cloning and expression of a trypanastigote-specific 85-kilodalton surface antigen gene from Trypanosoma cruzi.";
RL Mol. Biochem. Parasitol. 37:57-64(1989).
CC -!- FUNCTION: Implicated in attachment and penetration of host cells, possibly via a neuraminidase activity.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- SIMILARITY: Contains 2 BNR repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.ebi.ac.uk/announcements or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; M64836; AAA30150.1; -.
DR EMBL; J04667; AAA03205.1; -.
DR PIR; S25236; S25236.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR_2.
DR PRINTS; PR01803; TCSIALIDASE.
KW Signal; Multi-gene family; GPI-anchor; Antigen; Glycoprotein; Repeat; Membrane; Lipoprotein.
FT SIGNAL 1 24
FT CHAIN 25 691 85 kDa SURFACE ANTIGEN
FT PROPEP 692 714 REMOVED IN THE MATURE FORM
FT LIPID 691 691 GPI-anchor amidated alanine (Potential).
FT DOMAIN 26 29 POLY-ALA.
FT DOMAIN 150 153 POLY-THR.
FT DOMAIN 702 706 POLY-LEU.
FT REPEAT 258 268 BNR 1.
FT REPEAT 302 313 BNR 2.
FT CARBOHYD 96 96 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 124 124 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 171 171 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 475 475 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 546 546 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 652 652 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 714 AA; 77875 MW; 67800792790DC9AF CRC64;

Query Match
Best Local Similarity 23.7%; Score 832; DB 1; Length 714;
Matches 233; Conservative 97; Mismatches 275; Indels 88; Gaps 25;

QY 27 RVELEKRSQKVPFEKGGKTVRVVHSFRLPALVNDVGMVAIADARYETSN-DNSLI-- 83
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
43 RFLFVFPQKT-VLLPRGGNSKKWDSPASPLVSAGVIAAFAGHLSKKKDKNSTEP 101
QY 84 --DTVAKYSVDDGETWETQIAIKNS---RASSVSRVVD-----PTVIVKGNKLV 128
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
102 SSDAVAWY-IDSAWENSTLVGEVKNSTWQAHTVLGVKDGKRPDVLAPFTTTTKNKVPL 160
```


Db 500 TKVKGQFQLPEDPSRAVSNIPDGNVRHISLHNFLLVASVIEEAPSGNTPLTAVLV 559

Qy 527 SSGGKLLGLSYDEKQWQPIY--GSTVPTFGSWMGKRYVIVLTMANKIGSVYIDGEP 584

Db 560 DAGPEYFMRISYADKNWMTMLKDEKPTESRPWEAGKHOVALMLQGNKASVYVVDGEL 619

Qy 585 LEGSGQTVVP-----DGRTPDISHYVGGYGRSDMPITISHVTVNNV 625

Db 620 L---GEEVPLTEKPLEIFAFCFGACKIDGDEESSPEIKGKPR-----VTVNV 668

Qy 626 LLYNRLNAEERTL 640

Db 669 FLYNRLNSTEMRAI 683

RESULT 4

NANH CLOSO

ID NANH CLOSO STANDARD; PRT; 404 AA.

AC P15498;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Sialidase precursor (EC 3.2.1.18) (Neuraminidase).

OS Clostridium sordellii.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

OX NCBI_TaxID=1505;

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-45.

RC STRAIN=ATCC 9714 / NCBI 10717;

RX MEDLINE=90132537; PubMed=2693593;

RA Rothe B., Roggendorf P., Frank R., Bloeker H., Schauer R.;

RT "Cloning, sequencing and expression of a sialidase gene from Clostridium sordellii G12.";

RL J. Gen. Microbiol. 135:3087-3096 (1989).

CC -!- FUNCTION: SIALIDASES HAVE BEEN SUGGESTED TO BE PATHOGENIC FACTORS IN MICROBIAL INFECTIONS.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-, alpha-(2->8)-glycosidic linkages of terminal sialic residues in oligosaccharides, glycoproteins, glycolipids, colominic acid and synthetic substrates.

CC -!- SUBCELLULAR LOCATION: Periplasmic.

CC -!- PTM: IT IS POSSIBLE THAT THE SIALIDASE IS CLEAVED IN FRONT OF A CYSTEINE WITHIN THE LEADER PEPTIDE, FORMING A GLYCERIDE THIOESTER BOND WHICH LINKS THE PROTEIN TO THE MEMBRANE. A SECOND PROTEOLYTIC CLEAVAGE RELEASES THE MATURE EXTRACELLULAR PROTEIN.

CC -!- SIMILARITY: Belongs to family 33 of glycosyl hydrolases.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; M31584; AAA23280.1; --

DR PIR; A37234; A37234.

DR HSSP; P29768; 2SIL.

DR InterPro; IPR002860; GH_BNR.

DR InterPro; IPR00437; Prok lipoprot S.

DR Pfam; PF02012; BNR; 5.

DR SIGNAL 1 27

DR PROSITE; PS00133; PROKAR_LIPOPROTEIN; UNKNOWN 1.

KW Hydrolase; Glycosidase; Periplasmic; signal; Repeat.

FT CHAIN 28 404 SIALIDASE.

FT REPEAT 89 100 BNR 1.

FT REPEAT 158 169 BNR 2.

FT REPEAT 226 237 BNR 3.

FT REPEAT 273 284 BNR 4.

SQ SEQUENCE 404 AA; 44729 MW; 525B9DA90083AA6A CRC64;

Query Match 11.2%; Score 394.5; DB 1; Length 404;

Best Local Similarity 30.9%; Pred.No. 5.4e-21;

Matches 109; Conservative 67; Mismatches 136; Indels 41; Gaps 14;

Qy 54 FRIPALVNV-DGVMAIADARYETSDNLSLDTVAKYSVDDGETWETQIAIKNSRA-SSV 111

Db 54 FRIPSLQTLADGTMLAFSDIRYNGAEHVIDIGAAKSTDNGQTDWYKVTWMDRIDSTF 113

Qy 112 SRVVDPTIVKQ-NKLYVLVGSVNSRSYTWSHGDAR-DWDILLAVGEVTKTAGKITA 169

Db 114 SRVMDSTTVTDTGRIILTAGSNKNKNWASSTTSLSRSDWSVQMVYSDNGET- 166

Qy 170 STKMGSPVSLKEFFPAEMEGMHTNQ--FLGGAGVAIVASGNLNVYVPQVTV---NKKQV 224

Db 167 ---WSDKVDLTT-NKARIKNQPSNTIGWLAGVSGIVMSDGTIVMPIQIALRENNANN 222

Qy 225 SKIFYSEDEGKTWKGEGRSDFGCEPVALEWEGKLIINTRVDYRR-RLVYESSDMGNSW 283

Db 223 SSVIYKDNGETWTWGNKVPDPKTSNEMVIELDGLIMSSRNDGKNYASYSYDMGSTM 282

Qy 284 VEAVGTLSRVMGP----SPKSNQPGSQSFATVTI-EGMRVWMLFTHPLNFKRWLRDLN 338

Db 283 -----EVDPLHNKISTGNSCGQSFIVKTAQGHRLGFISAPKNTKGGYVRDNT 334

Qy 339 LMLTD-----NORIYVNGQVSGIDNSA---YSSVLYKDDKLYCLHEINSNEVY 384

Db 335 VTMDFDLDSKIRELCSPYPEDGNSGGYSCLSFNDGKLSILYEANGNIEY 387

RESULT 5

NANH CLOPE

ID NANH CLOPE STANDARD; PRT; 382 AA.

AC P10481;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Sialidase (EC 3.2.1.18) (Neuraminidase).

GN NANH.

OS Clostridium perfringens.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

OX NCBI_TaxID=1502;

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 8-27.

RC STRAIN=A99;

RX MEDLINE=89005674; PubMed=2901987;

RA Roggendorf P., Rothe B., Lottspeich F., Schauer R.;

RT "Cloning and sequencing of a Clostridium perfringens sialidase gene.";

RL FEBS Lett. 238:31-34 (1988).

CC -!- FUNCTION: SIALIDASES HAVE BEEN SUGGESTED TO BE PATHOGENIC FACTORS IN MICROBIAL INFECTIONS.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-, alpha-(2->8)-glycosidic linkages of terminal sialic residues in oligosaccharides, glycoproteins, glycolipids, colominic acid and synthetic substrates.

CC -!- SUBCELLULAR LOCATION: Periplasmic.

CC -!- SIMILARITY: Belongs to family 33 of glycosyl hydrolases.

CC -!- SIMILARITY: Contains 4 BNR repeats.

CC -!- DATABASE: NAME=Worthington enzyme manual;

CC WWW="http://www.worthington-biochem.com/NEUP/".

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; Y00963; CAA68780.1; --

DR PIR; S01339; S01339.


```
DR HSP; P29768; 2SIL.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 4.
DR PRINTS; PR01803; TCSIALIDASE.
KW Hydroxylase; Glycosidase; Periplasmic; Repeat.
FT REPEAT 71 82 BNR 1.
FT REPEAT 140 151 BNR 2.
FT REPEAT 208 219 BNR 3.
FT REPEAT 255 266 BNR 4.
SQ SEQUENCE 382 AA; 42813 MW; D50857BB50A4E886 CRC64;

Query Match 11.0%; Score 388; DB 1; Length 382;
Best Local Similarity 30.4%; Pred. No. 1.5e-20;
Matches 109; Conservative 72; Mismatches 138; Indels 40; Gaps 16;

QY 54 FRLP--ALVNDGVMAIDARVETSDNSLDITVAKYSVDDGETWETQIAKNSRA-SS 110
DB 36 FRIPNTQLLN-DGITLTFSDIRYNGPDHAYIDIASARSTDFGKTWSYNIAMKNNRIDST 94
QY 111 VSRVVD-PTVIVKGNKLYLVLSYNSRSY-WTSHGDARDWDILLAVGEVTKSTAGGKIT 168
DB 95 YSRVMDSTVITNTGRIILLAGSWNTNGNWMATITSTERSDWSQMIYSD-----D 144
QY 169 ASIKWSPVSL-KEFPFAEMEGMHTNQFLGGAGVAIVASGNLIVYPQVT---NKKKQVF 224
DB 145 NGLTWSNKIDLTAKDSSKXNQPSNTIGLGGVSGIVMDGTTVMPAQISLRENNENNY 204
QY 225 SKIFYSDEGKTKFEGRSDFCCSEPVALEWEGKLIINTRVDYR-RLVYESSDNGNSW 283
DB 205 SLIIYSKDNGETWTKNGKVPNSNTSENWVIELDGLIMSTRIDYSGYRAAYISHDLGTW 264
QY 284 VEAVGTLRVWGPSKSNQPGSSP-TAVTIEGRVMLFTHPLNFKGRWLRDLNLWLT 342
DB 265 -EIEYELN---GKLLTGKSGCGGCFIKATTSNGHIGLISAPKNTKGEYIRDNIAYMI 320
QY 343 D-NQRIYVQVSI-----GDE-NSAYSVLYKDDKLYCLHEINSN-----EVSIV 387
DB 321 DFDLSKGVQEIPIYPEDGNKLGCGYCLSPKNNHLGIYVEANGNIEYQDLTPPYSLI 379

RESULT 6
NANH_SALTY STANDARD; PRT; 381 AA.
AC P29768;
DT 01-APR-1993 (Rel. 25, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sialidase (EC 3.2.1.18) (Neuraminidase) (NANase) (N-acylneuraminase)
DE glycosylase (STNA).
GN NANH OR STM0928.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
ON NCBI_TaxID=602;
RX SEQUENCE FROM N.A., AND SEQUENCE OF 1-39.
RC STRAIN=LT2;
RX MEDLINE=92292952; PubMed=1602967;
RA Hoyer L.L., Hamilton A.C., Steenberg S.M., Vimr E.R.;
RT "Cloning, sequencing and distribution of the Salmonella typhimurium
RT LT2 sialidase gene, nanH, provides evidence for interspecies gene
RT transfer.";
RL Mol. Microbiol. 6:873-884(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
```

```
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP CHARACTERIZATION, AND X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RC STRAIN=LT2;
RX MEDLINE=92389334; PubMed=1518058;
RA Taylor G.L., Vimr E.R., Garman E.F., Laver W.G.;
RT "Purification, crystallization and preliminary crystallographic study
RT of neuraminidase from Vibrio cholerae and Salmonella typhimurium
RT LT2.";
RL J. Mol. Biol. 226:1287-1290(1992).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS), AND REVISIONS TO C-TERMINUS.
RC STRAIN=LT2;
RX MEDLINE=94052190; PubMed=8234325;
RA Crennell S.J., Garman E.F., Laver W.G., Vimr E.R., Taylor G.L.;
RT "Crystal structure of a bacterial sialidase (from Salmonella
RT typhimurium LT2) shows the same fold as an influenza virus
RT neuraminidase.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:9852-9856(1993).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
RC STRAIN=LT2;
RX MEDLINE=96228252; PubMed=8656428;
RA Crennell S.J., Garman E.F., Philippon C., Vasella A., Laver W.G.,
RA Vimr E.R., Taylor G.L.;
RT "The structures of Salmonella typhimurium LT2 neuraminidase and its
RT complexes with three inhibitors at high resolution.";
RL J. Mol. Biol. 259:264-280(1996).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.05 ANGSTROMS).
RA Garman E.F., Wouters J., Schneider T.R., Vimr E.R., Laver W.G.,
RA Sheldrick G.M.;
RL Submitted (JUL-1998) to the PDB data bank.
CC -!- FUNCTION: CLEAVES THE TERMINAL SIALIC ACID (N-ACETYL NEURAMINIC
CC ACID) FROM CARBOHYDRATE CHAINS IN GLYCOPROTEINS PROVIDING FREE
CC SIALIC ACID WHICH CAN BE USED AS CARBON AND ENERGY SOURCES.
CC SIALIDASES HAVE BEEN SUGGESTED TO BE PATHOGENIC FACTORS IN
CC MICROBIAL INFECTIONS.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2-3)-, alpha-(2-6)-,
CC alpha-(2-8)-glycosidic linkages of terminal sialic residues in
CC oligosaccharides, glycoproteins, glycolipids, colominic acid and
CC synthetic substrates.
CC -!- SUBUNIT: Monomer.
CC -!- SIMILARITY: Belongs to family 33 of glycosyl hydrolases.
CC -!- SIMILARITY: Contains 4 BNR repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M55342; AAA27168.1; -.
DR EMBL; AB008737; AAL19864.1; ALT_INIT.
DR PDB; 2SIL; 31-AUG-94.
DR PDB; 3SIL; 13-JAN-99.
DR PDB; 2SIM; 30-NOV-94.
DR PDB; 1DIL; 07-DEC-96.
DR PDB; 1DIM; 07-DEC-96.
DR StyGene; SG10244; nanH.
DR InterPro; IPR002860; GH_BNR.
DR InterPro; IPR008377; Sialidase_trypan.
DR Pfam; PF02012; BNR; 4.
DR PRINTS; PR01803; TCSIALIDASE.
KW Glycosidase; Hydrolase; Repeat; 3D-structure; Complete proteome.
FT INIT MET 0 0
FT REPEAT 70 81 BNR 1.
FT REPEAT 144 155 BNR 2.
FT REPEAT 209 219 BNR 3.
```

```
FT REPEAT 253 264 BNR 4.
FT DISULFID 41 102
FT ACT_SITE 36 36
FT ACT_SITE 245 245
FT ACT_SITE 308 308
FT CONFLICT 328 328
FT STRAND 4 8
FT TURN 10 11
FT STRAND 16 16
FT TURN 18 19
FT STRAND 22 22
FT STRAND 27 27
FT HELIX 28 30
FT STRAND 31 31
FT TURN 34 42
FT STRAND 44 45
FT STRAND 48 55
FT STRAND 65 72
FT STRAND 80 85
FT TURN 92 94
FT STRAND 96 107
FT TURN 108 109
FT STRAND 110 120
FT HELIX 127 129
FT TURN 137 138
FT STRAND 140 146
FT TURN 149 150
FT HELIX 154 156
FT STRAND 160 167
FT STRAND 170 175
FT STRAND 178 179
FT STRAND 181 182
FT TURN 184 185
FT STRAND 188 196
FT TURN 198 199
FT STRAND 204 211
FT STRAND 217 218
FT STRAND 224 225
FT TURN 227 228
FT STRAND 231 236
FT TURN 237 238
FT STRAND 239 244
FT STRAND 249 249
FT STRAND 251 251
FT STRAND 253 255
FT STRAND 263 264
FT TURN 266 270
FT STRAND 273 273
FT TURN 276 277
FT STRAND 282 288
FT TURN 289 290
FT STRAND 291 299
FT TURN 302 303
FT TURN 306 307
FT STRAND 311 316
FT TURN 318 320
FT STRAND 323 330
FT TURN 336 337
FT STRAND 342 349
FT TURN 350 351
FT STRAND 352 361
FT TURN 362 363
FT STRAND 364 369
FT HELIX 371 373
FT TURN 374 378
FT TURN 379 379
SQ SEQUENCE 381 AA; 41942 MW; 584D54DE142F2165 CRC64;

Query Match
Best Local Similarity 9.8%; Score 344; DB 1; Length 381;
Matches 116; Conservative 63; Mismatches 154; Indels 78; Gaps 18;

30 LFKRSSKVPPEKG-----GKTVRVHSLPALVND-GVMVAIDARYETSDNSLID 84
QY
```

```
DR InterPro; IPR008377; Sialidase trypan.
KW PRINTS; PRO1803; TCSIALIDASE.
FT NON_TER 1
SQ SEQUENCE 175 AA; 19552 MW; E90DD13274B75B8F CRC64;

Query Match      8.3%; Score 291.5; DB 1; Length 175;
Best Local Similarity 39.0%; Pred. No. 4.3e-14;
Matches 69; Conservative 26; Mismatches 75; Indels 7; Gaps 4;

QY 458 CUNASTANAEVFNGLKFPAGVGGALVPVSOQONQRYHFANHAFTLVASVTHVEVPSVA 517
DB 2 CLNATVNRATKVKDGFQLTPEPSGVMPVNPIDYKRGHVFNLNFTLVASVTIEEAPSGN 61

QY 518 SPILGASLDSGCKLGLSLYSDEKHWOPIY--GSTPTVPTGSGWEMGRYHVVLWTANKI 575
DB 2 TPLLIIVLANTETHWRIILYTADNKWMTLMDKEKPTTSGTWEPKKEHQVALMLQGNK 121

QY 576 GSYVIDEPLGSGQTVP-DGRTP-DISHFYVGGYGRSDMPTTISHVTVNNVLLYNR 630
DB 122 ASYVVDGELL--GEEVEPLTGKPLELFAFCGACGEENPQSBHVTVTNVFLYNR 175

RESULT 8
NANA_STRPN STANDARD; PRT; 1035 AA.
AC Q59959; Q54722;
DT 15-DEC-1998 (Rel. 37, Created)
DT 28-FEB-2003 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sialidase A precursor (EC 3.2.1.18) (Neuraminidase A).
GN NANA OR SPRI536.
OS Streptococcus pneumoniae, and
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313, 171101;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R36A / NCTC 10319;
RX MEDLINE=94341870; PubMed=8063384;
RA Camara M., Boulnois G.J., Andrew P.W., Mitchell T.J.;
RT "A neuraminidase from Streptococcus pneumoniae has the features of a
RL surface protein.";
RL Infect. Immun. 62:3688-3695 (1994).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-255 / R6;
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,
RA Norris F.A., O'Garra M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Balz R.H., Jaekunas S.R., Rostock P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717 (2001).
[3]
RP SEQUENCE OF 882-1035 FROM N.A.
RC STRAIN=Serotype 6;
RX MEDLINE=96326329; PubMed=8759848;
RA Berry A.M., Lock R.A., Paton J.C.;
RT "Cloning and characterization of nanB, a second Streptococcus
RT pneumoniae neuraminidase gene, and purification of the NanB enzyme
RT from recombinant Escherichia coli.";
RL J. Bacteriol. 178:4854-4860 (1996).
CC -1- CATABOLIC ACTIVITY: Hydrolysis of alpha-(2-3)-, alpha-(2-6)-,
CC alpha-(2-8)-glycosidic linkages of terminal sialic residues in
CC oligosaccharides, glycoproteins, glycolipids, colominic acid and
CC synthetic substrates.
```

```
CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (Potential).
CC -1- SIMILARITY: Belongs to family 33 of glycosyl hydrolases.
CC -1- SIMILARITY: Contains 4 BNR repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X72967; CAAS1473.1; -.
CC EMBL; AE008522; AAL00340.1; -.
CC EMBL; U43526; AAC44391.1; -.
CC PIR; G98063; G98063.
CC PIR; T30287; T30287.
CC HSP; Q02834; IEUR.
CC InterPro; IPR008985; ConA like_lect_gl.
CC InterPro; IPR002860; GH_BNR.
CC InterPro; IPR004124; Glyco hydro 33N.
CC InterPro; IPR005877; Gpos_Y8IRK_.
CC InterPro; IPR001899; Gram_pos_anchor.
CC InterPro; IPR001791; Laminin_G.
CC Pfam; PF02012; BNR; 4.
CC Pfam; PF02973; sialidase_N; 1.
CC Pfam; PF04650; Y8IRK_signal; 1.
CC SMART; SM00282; LamG_1.
CC TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
CC TIGRFAMs; TIGR01168; Y8IRK_signal; 1.
CC PROSITE; PSS0847; GRAM_POS_ANCHORING; 1.
CC Hydrolyase; Glycosidase; Cell wall; Peptidoglycan-anchor; Repeat;
KW Signal; Complete proteome.
FT SIGNAL 1 53 POTENTIAL.
FT CHAIN 54 1006 SIALIDASE A.
FT PROPEP 1007 1035 REMOVED BY SORTASE (POTENTIAL).
FT SITE 1003 1007 LPXTG SORTING SIGNAL (POTENTIAL).
FT MOD_RES 1006 1006 AMIDE-LINKING TO CELL WALL (POTENTIAL).
SQ SEQUENCE 1035 AA; 114741 MW; C5B8A2D7A12E12F3 CRC64;

Query Match      6.4%; Score 226.5; DB 1; Length 1035;
Best Local Similarity 20.9%; Pred. No. 2.7e-08;
Matches 139; Conservative 78; Mismatches 212; Indels 237; Gaps 32;

QY 27 RVLEFKRQ--SSKVP-----FEKG--GKTVRVVHSFRLPALVVD-GVMVAIA 70
DB 304 RSLFRSRLDLEKLPDEGALTEKIDIFESGRNGKPKDGKISYRIPALLKTDKGTLLAGA 363
QY 71 DARYETSDNNSLIDTVAKYSVDGETWETQIAIKNSRASSVSRVDPD-----VIV 121
DB 364 DERRLHSSDWGDIGWIRRSSENGKTWDRVTITNLRDN--PKASDPSIGSPVNMIDMLV 421
QY 122 -----KG-----NKLIVLV-----CSY--NSRSY 139
DB 422 QDPETKRIFSYDMFPEGKIFGMSQKEAYKKIDGKTYQILYREGKGYTIRENGTV 481
QY 140 WTSHGDARDWDILL-----AVGEV---TKSTAGKITASI----- 171
DB 482 YTPDGKATDYRVVVDVVKPAYSKDGLYKGNQLGNLYFTNTKTSFPRIADKSYLWMSYS 541
QY 172 -----KWGSPVSLKFPFPAEMEGMHTNQFLG-GAGVAIVASN-----GNLYVPQVVTNKKK 221
DB 542 DDGKTSAPQDITPVMKADW-----MKFLGVGPGTGLVLRNGPHKGRILLIPVVTNNVS 596
QY 222 QV-----FSKIFYSEDGKTKWFGESRSDPCSPVALEWEGKLIINTRDYRR----- 271
DB 597 HLNGSQSSRIIYSDHGKTHWAGEVND-----NRQVDGQKIHSSHTMNNRAQNTST 649
QY 272 -----LVYESDMSGNSVVEAVGVTLSRVNGSPSKNQPGSQSFTA 311
DB 650 VVOLNNGDVKLFMRGLTGDILQVATSKGGVTWEKDKRY-----PQVKDYYVQMSAIH 702
```

Qy 312 VTTEGRVMFLFTH---PLNFKGRLDRLN-----LWLTDN----ORIYNVGQVSTIGDE 358
Db :::|::| || || || || || || || ||
703 TMEHGEKEYITLSNAGPKRENGMWHLARVEENGELTWLKINPIQGFAPN-----SLQEL 758
Qy 359 NSAYSSVLVKDDKLYCLCHEINSNEVYSLVPARLVGELRIIKSVLQSWKNWDSSLSSICTP 418
Db :::|::| || || || || || || || ||
759 GNGEYCILYE-----HTEKGQNAYTISPRKF-----NWDFLSKDLISP 796
Qy 419 ADPAASSSSRGCGPATVTTVGLVGLFSHSATKTWEEDAYRCVNASTA-----N 465
Db :::|::| || || || || || || || ||
797 TE-AKVGRTEMGK-----GVIGLEFDSEVLNVKAPTQLANGKYARFMQTDTKTLFT 850
Qy 466 ASERVENGLKFAGVGGAL-----WPVS-----QQGNQRHYHPANHAFTLVASVTIHEYP 514
Db :::|::| || || || || || || || ||
851 VDSEDWGOKVTGLAEAGATESMNLFPVSAGTKLSNGMG-----SEAAVHEVP 898
Qy 515 SVASPL 520
Db ||
899 EYTGPL 904
Db

```

RESULT 9
ID 8512 TRYCR
ID 8512 TRYCR STANDARD; PRT; 240 AA.
AC P18270;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sialidase 85-1.2 (EC 3.2.1.18) (Neuraminidase) (NA) (Major 85 kDa
DE surface antigen) (SA85-1.2 protein) (Fragment).
DE SA85-1.2.
DE Trypanosoma cruzi.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
NCBI_TaxID=5693;
RN [1]
SEQUENCE FROM N.A.
SEQUENCE CL;
STRAIN=CL;
MEDLINE=90324879; PubMed=1695668;
Kahn S., van Voorhis W., Eisen H.;
"the major 85-kD surface antigen of the mammalian form of Trypanosoma
cruzi is encoded by a large heterogeneous family of simultaneously
expressed genes.";
J. Exp. Med. 172:589-597(1990).
-1- FUNCTION: Developmentally regulated neuraminidase implicated in
parasite invasion of cells. May contribute to the pathology during
T.cruzi infection by cleaving sialic acid from cells of the immune
system.
-1- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-,
alpha-(2->8)-glycosidic linkages of terminal sialic residues in
oligosaccharides, glycoproteins, glycolipids, colominic acid and
synthetic substrates.
-1- DEVELOPMENTAL STAGE: Mammalian stage of parasite.
-1- MISCELLANEOUS: The parasite mammalian stage surface antigen
exhibits extensive antigenic diversity.
-1- SIMILARITY: Belongs to family 33 of glycosyl hydrolases.
-----
THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X53546; CAA37618.1; --
PIR; S11293; S11293.
InterPro; IPR006377; Sialidase_trypan.
PRINTS; PK01803; TCSIALIDASE.
SIALIDASE; Glycosidase; Repeat; Multigene family; Antigen.
NON TER
1
SEQUENCE 240 AA; 26600 MW; 8B73A9F7EE19ED9C CRC64;
Query Match 5.9%; Score 208; DB 1; Length 240;

```

Query Match 5.9%; Score 208; DB 1; Length 240;

Best Local Similarity	32.3%;	Pred. No. 7.2e-08;	
Matches	61;	Conservative	15; Mismatches 63; Indels 50; Gaps 4;
QY	490	QGNQRYIFANHAFTLVASVTITHVEVPSVASPLLGASLDSSGKKLGLGSHDEKHOWQPIYG	549
DB	4	GDNVRHVLNHNFTLVASVTIEEAPSEKTPLLTALLGDAEPFPMRLSYTADNKKWETISK	63
QY	550	STPVPTPGS--WEMGKEYHVLTWMANKIGSVYIDGEP-	585
DB	64	GDKKLTETSPWPVKEHQVALMLQGNKASVYIDGESLGEAEPLVTETPLEPFGFCFGAC	123
QY	586	-----EGSGQTVPDGRTPDISHFYGVGYGRSDMPTTISHVTNNVLNLRQ	631
DB	124	DFDDDDGGDDDEEDSQEESPKESSPE-----KIGKK-----PHVTVTNNVFLYNRP	171
QY	632	LNABEINTL	640
DB	172	LNPTENRAI	180

RESULT 10
NANH CLOSE

ID	NANH_CLOSE	STANDARD;	PRT;	1014 AA.
AC	P29767;			
DT	01-APR-1993 (Rel. 25, Created)			
DT	01-APR-1993 (Rel. 25, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Sialidase precursor (EC 3.2.1.18) (Neuraminidase).			
OS	Clostridium septicum.			
OS	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;			
OC	Clostridium.			
NCBI_TaxID=1504;				
OX	[1]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=NC 0054714;			
RC	STRAIN=91238693; PubMed=2034213;			
RX	Rothe B., Rothe B., Roggenstein P., Schauer R.;			
RA	"The sialidase gene from Clostridium septicum: cloning, sequencing,			
RT	expression in Escherichia coli and identification of conserved			
RT	residues in sialidases and other proteins.";			
RT	Mol. Genet. 226:190-197(1991).			
RL	-1- FUNCTION: SIALIDASES HAVE BEEN SUGGESTED TO BE PATHOGENIC FACTORS			
CCC	IN MICROBIAL INFECTIONS.			
CCC	-1- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-,			
CCC	alpha-(2->8)-glycosidic linkages of terminal sialic residues in,			
CCC	oligosaccharides, glycoproteins, glycolipids, colominic acid and			
CCC	synthetic substrates.			
CCC	-1- SUBCELLULAR LOCATION: Periplasmic.			
CCC	-1- SIMILARITY: Belongs to family 33 of glycosyl hydrolases.			
CCC	-1- SIMILARITY: Contains 4 BNR repeats.			
CCC	-1- SIMILARITY: Contains 1 FS/8 type C domain.			
CCC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CCC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CCC	the European Bioinformatics Institute. There are no restrictions on its			
CCC	use by non-profit institutions as long as its content is in no way			
CCC	modified and this statement is not removed. Usage by and for commercial			
CCC	entities requires a license agreement (See http://www.eib-sib.ch/announcement/			
CCC	or send an email to license@sib-sib.ch).			
CCC	-----			
CCC	EMBL; X63266; CAA44916.1; -.			
CCC	PIR; S15994; NMCLASS.			
CCC	HSP; Q02834; LEUR.			
CCC	InterPro; IPR008985; ConA like lec_gl.			
CCC	InterPro; IPR00421; FA58 C.			
CCC	InterPro; IPR008979; Gal Bind like.			
CCC	InterPro; IPR002860; GH_BNR.			
CCC	InterPro; IPR004124; Glyco_hydro_33N.			
CCC	InterPro; IPR00212; BNR; 4.			
CCC	Pfam; PF02973; sialidase_N; 1.			
CCC	PROSITE; PS50022; FAS8C_3; 1.			
CCC	Hydrolase; Glycosidase; Periplasmic; Signal; Repeat.			
CCC	SIGNAL			
CCC	1 26			
CCC	POTENTIAL.			

FT CHAIN 27 1014 SIALIDASE.
FT DOMAIN 39 186 F5/8 TYPE C.
FT REPEAT 431 442 BNR 1.
FT REPEAT 563 574 BNR 2.
FT REPEAT 627 638 BNR 3.
FT REPEAT 700 711 BNR 4.
SQ SEQUENCE 1014 AA; 110652 MW; C4F49233473A2FAD CRC64;

Query Match 5.5%; Score 195; DB 1; Length 1014;
Best Local Similarity 20.0%; Pred. No. 5e-06;
Matches 154; Conservative 117; Mismatches 235; Indels 264; Gaps 42;

QY 16 PRGSHAPGSSRVLP-----KQSSKVPPEKGVKIVRV----- 50
Db 331 PGSGNEYNFTGEIDFFELXSKPLADRYLKERTGETTSKDLFPFEGAVKTEPVDIPTFGEL 390
QY 51 -VHSFRLPALVNV-DGVMAIADARVETSND-NSLIDTVAKYSDVDETW-ETQIAIKNS 106
Db 391 GSNFRIPALYTTKOGTVLASIDVRKGGHDAPNNIDTIGIKRSTGGVWDEGKIILDYP 450
QY 107 RASSV-----SRVVDPTVIVKGNKLYLVVGSYNSSR 137
Db 451 GASSAIDTSLQDDETRIFLIIVTHFAEGYFGNCKTSGYVEIEGKRYLKLIGANDTIY 510
QY 138 S-----YWTSHGDARDWI-----LLAVGEVTKSTAGGKITASI----- 171
Db 511 TVREGVYDSNGEATYTVDDNNELYENGRIGNVLLSNSPLKVMGTGTFSLIYSDDDGQ 570
QY 172 KMGSPVSLKEFFPAEMEGMHTNQFLG--GAGVAIVASN--GNLVVPQVQVTKKKQVFSK 226
Db 571 TWSDPIDLNKEVKTDM-----MRPLGTGPGKHQIKTGRYAGRLPLFVILTNASGQSSA 625
QY 227 IFYSEDEGKTKFGEGRSDFGCEPVALEWEGKLIINTRVYRRRLVYESSDMGNSWVEA 286
Db 626 VIYSDNGATWNETATD-----GRLM-----DNGDRASA--ET 658
QY 287 VGTLSRVWGPSPKSNQPGSQSSFTATIEGMVWLFTHPLNFKRWLRDLRLNLWLTNQR 346
Db 659 ITT-----NTSGGVGQUTEQCVVEMP-----NGQLKMFMENTGG 692
QY 347 IYNVGOVSTG---DENSAYSSVLYKDDKL---YC-LHEINSNEVSYLVFARLVGELRIIK 399
Db 693 --NSGRVIATSPDGATWEDDVVRDENIKBPYCOLSVINYSQ-----KIDGKDAILF 743
QY 400 SVLQSKWQNDSHLSSICTPADPAASSRSGCGPAVTVGLV---GFLSHSATK--TEWE- 453
Db 744 AI-----PDANYPNRVNG--TVRVGLITENGSEYENGEPRYDIEMRY 782
QY 454 -----DAYRCVNASTANAERPVN---GLKFAGVGGGALWPVSOQGNQVRVHFANHAF 502
Db 783 NKVVPAGTYGYSCLS-----EMPNGEIGLFYEGRG-----SRQMSFTRMNI 823
QY 503 TLVASVTIHEVPSVASPLGASLDSS-----GGKLLGLSYDEKHQNPYIGSTPTPTGS 558
Db 824 DYLLKADLLQVPAANTKSYTNSNNIYDPGKISLNVTFD---QTVSLIGDRITTA--- 877
QY 559 WEMGKRYHYVLTWANKIGSVYIDGEPLEGSGQVWPDGRTP-DISHFTYGGY-----GRS 612
Db 878 -DIGGK-EVLLTLANSKG-----GSEYTF--EGTVPADISN---GNVTITIKGS 920
QY 613 DMPTISHVTNNVLLYNRQLN-----ABEIRTL--FLSQDLIGTEAHM 653
Db 921 GLKIVN--VVKVTDITEDRNTGLNVQGVBEVQSDVKTLQLQDLVDSNLSNL 968

RESULT 11
ID NANH_BACFR STANDARD; PRT; 266 AA.
AC P31206;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sialidase (EC 3.2.1.18) (Neuraminidase) (Fragment).

GN NANH.
OC Bacteroides fragilis.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=817;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TAL2480;
RA MEDLINE=90336922; PubMed=2158974;
RX Russo T.A., Thompson J.S., Godoy V.G., Malmay M.H.;
RT "Cloning and expression of the Bacteroides fragilis TAL2480
neuraminidase gene, nanH, in Escherichia coli.";
RL J. Bacteriol. 172:2594-2600(1990).
CC -!- FUNCTION: SIALIDASES HAVE BEEN SUGGESTED TO BE PATHOGENIC FACTORS
CC IN MICROBIAL INFECTIONS.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-,
CC alpha-(2->8)-glycosidic linkages of terminal sialic residues in
CC oligosaccharides, glycoproteins, glycolipids, colominic acid and
CC synthetic substrates.
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -!- SIMILARITY: Belongs to family 33 of glycosyl hydrolases.
CC -!- SIMILARITY: Contains 5 BNR repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; M31663; AAA22912.1; -;
DR HSSP; Q02834; 1EUR.
DR InterPro; IPR002860; GH_BNR.
DR InterPro; IPR008377; Sialidase_trypan.
DR Pfam; PF02012; BNR; 5.
DR PRINTS; PR01803; TCSIALIDASE.
DR Hydrolyase; Glycosidase; Periplasmic; Repeat.
KW NON_TER 1 1
FT REPEAT 5 16 BNR 1.
FT REPEAT 84 95 BNR 2.
FT REPEAT 191 202 BNR 3.
FT NON_TER 266 266
SQ SEQUENCE 266 AA; 29035 MW; 22ADC720F97DA6BB CRC64;

Query Match 5.4%; Score 191.5; DB 1; Length 266;
Best Local Similarity 26.0%; Pred. No. 1.3e-06;
Matches 69; Conservative 46; Mismatches 103; Indels 47; Gaps 12;

QY 90 SVDDGETWETQ-----IAIKNSRASSVSRVDPVTIV--KGNKLYVLVG---SYNSRSRY 139
Db 7 STDGKTKWEKMRPLPAFGFGLPAGQNGVGDPSILVDTKTNNVWVVAWTHGMGNQRAW 66
QY 140 WSHGDARDWD-----ILLAVGEVTKSTAGKITASIKGSPVSLKEFFPAEMEGMHTNQF 195
Db 67 WSSH-PGMDMNHNTAQVLVA-----KSTDGK-----TWSAPINITE---QVKDPSWYFL 111
QY 196 LGAGAGVAIVASGNLVYPVQVTKKKQVFSKIFYSDEGKTKWFGSGRSDFGCSEPVALE 255
Db 112 LQPGRGITWSDGTLVFPPTQFDISTRPVAGINYSKDGKNWKNHNYARTNTTEAQVAEA 171
QY 256 WEGKLIINTRVDYR--RRLVYESSDMGNSVWEAVGTLRSVWGPSPKSNQPGSQSSFTAVT 313
Db 172 EPGVLMNMR-DNRGGRAVAITKDLGKTWTE-----HSSSRKALPESVCMASLIS 221
QY 314 IE-----GMRVWLFTHPLNFKGRW 332
Db 222 VKAKNVLGKOLLIFSNPNTTKGRY 246

RESULT 12
ID NANH_MICVI STANDARD; PRT; 647 AA.

199	YLAGSGDEISLFNKAISQEVSTIPLSNPQOLIFQSGDSTQ--ANYFRIPTYLTLSGR	256
66	MVAIADARYETSDN-SLIDTVAKYSVDGETWETQI-----AIKN	105
257	VLSGIDARYGTHDSKSKINIATSYSDONGKTWSEPIFAMKFNDEYEQLYWPRDNKLN	316
106	SRASSVSRVDPPIV--VKGNKLYL-----VGSYNSRS	146
317	SOISGSGASFIDSSIVEDKSGKTIILLADVMPAGIANNANKADSGFEINGHYLYLKK	376
147	RDMDILLAYGE-----VTKST-----AGKKTASIKK-----	173
377	GDNDFRYTVRENGVVYNTTNKPTNYTINDKYVELEGKSLTVEQYSVDFSGSLRERIN	436
174	GSPVSLKEPF-----PAEWECHTNO-----FLG-----GAGV	201
437	GKQVPMNVFYKOSLFKVTPTNYIAMTTSQNRGESWEQFKLLPPLFGEKHNGTVLCPGOGL	496
202	AIVASN-----GNLVYPVQVTKKKQVFSKIFYSEDEGKTKWFGESRDF--GCSE	250
497	ALKSNRLIFATVYTSGLTYLI-----SDSGQTKWKSASPIFFKNATF	541
251	PVALWEGLIINTRVDYR--RLVTESS--DMGNSWVEAVGTLSRVWGPSPKSNQPSQ	306
542	AQWELURDGV-----RTFPRITTTGKIAVMTSRDSETW-----SKVSYIDGIQOTS	592
307	SSFTAVT--IEGMRVMLFTHPLNFKGRWLDRNLMLTNQRIYNNGVQVSGIDENSAYSS	364
593	VSAIKYSQLIDGKEAVILLSTPNSR-----KCGVLUVLKNKEDS	635
365	VLYKDDKLYCLHEINSNEVSLVPARLVGELRIIKSVLQSWKNWDSHJSSI	415
636	IDWKTH--YDIDLPSYGVAYSATLPELNHHIGVLPEKDYSWRNEHLHLSNT	684

RESULT 14
APU THEFT

ID	APU	THETU	STANDARD;	PRT;	1861	AA.
AC	P38536;					
AD	01-OCT-1994	(Rel. 30, Created)				
AT	01-NOV-1995	(Rel. 32, Last sequence update)				
DT	10-OCT-2003	(Rel. 42, Last annotation update)				
DT	Amypollulanase precursor	(Alpha-amylase/pullulanase)				(Pullulanase
DE	type II)	[Includes: Alpha-amylase (EC 3.2.1.1) (1,4-alpha-D-glucan				glucanohydrolase); Pullulanase (EC 3.2.1.41) (1,4-alpha-D-glucan
DE	glucanohydrolase);	Pullulanase (EC 3.2.1.41) (1,4-alpha-D-glucan				glucanohydrolase) (Alpha-dextrin endo-1,6-alpha-glucosidase)].
DE	AMYNB.					
GN	Thermoanaerobacter thermosulfurogenes	(Clostridium				
OS	thermosulfurogenes)					
OS	Thermosulfurogenes)					
OC	Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;					
OC	Thermoanaerobacteriaceae; Thermoanaerobacterium.					
OX	NCBI_TaxID=33950;					
RRN	[1]					
RRP	SEQUENCE FROM N.A.					
RRRC	STRAIN=DSM 3896 / EMI;					
RRRC	MEDLINE=94252998; PubMed=8195085;					
RRRA	Matuschek M., Burchhardt G., Sahn K., Bahl H.;					
RRRT	"Pullulanase of Thermoanaerobacterium thermosulfurogenes EMI					
RRRT	(Clostridium thermosulfurogenes): molecular analysis of the gene,					
RRRT	composite structure of the enzyme, and a common model for its					
RRRT	attachment to the cell surface.";					
RRRL	J. Bacteriol. 176:3295-3302(1994).					
CC	-I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic					
CC	linkages in oligosaccharides and polysaccharides.					
CC	-I- CATALYTIC ACTIVITY: Hydrolysis of (1->6)-alpha-D-glucosidic					
CC	linkages in pullulan and in amylopectin and glycogen, and the					
CC	alpha- and beta-limit dextrins in amylopectin and glycogen, and the					
CC	COFACTOR: Binds 1 calcium ion per subunit (By similarity).					
CC	-I- SUBCELLULAR LOCATION: CELL-BOUND. IT C-TERMINUS MAY SERVE AS AN					
CC	S-LAYER ANCHOR.					
CC	-I- PTM: Glycosylated.					
CC	-I- SIMILARITY: Belongs to family 13 of glycosyl hydrolases.					
CC	-I- SIMILARITY: Contains 2 fibronectin type III domains.					

Db 1119 TPKIGQWE-----YLMRP-----SDNQGDWI--TTSTLSFYVVPDLDLIKPTAPYL 1164
QY 301 NOPGSSSTAVTIEGMRVMLFTHPLNFKRWLRDLNLWLTQRIYVNVGVQVSGIDENS 360
Db 1165 NQPGTESS-----RVSLTWP-----STDNVCIYD-----1190
QY 361 AYSSVLYKDD-----KLYCLHEINSNEVSLVPAVLGRLRIKSVLQSKWNSHLSI 415
Db 1191 ----EYRSDGGTFNKIATV-----SNEVNYI-----DTSVINGVTVYKVVAVDLSPN 1236
QY 416 CTPADPAASRSGCGPATTVCLVGLFSLHSATKTEWEDAYRCVNASTANAERVPNGLKF 475
Db 1237 RTESNVVTIK-----PDVVPIKVI-----FNVTVDPY-----TPDAVNL 1270
QY 476 AGVGGGALVPVSOQQNQRYHPANHPTILVASVTHIEVPSVASPLLGASLDS-----527
Db 1271 AGTFPNATWDPSAQ-----QMTKIDNNTYSI--TILDEGQIEKYARGSWDKVEKDEYGN 1325
QY 528 --SGGKKLLGLSY-----DEKHQWQ--PIYGSTPTVPTTGSWEMGKRYHVVLTMANKI 575
Db 1326 EFASNRKVTIVNQGNEMTINDTVYRWDRDPIFIYSPSSN-----MTVDSNI 1372
QY 576 GSVYIDGCEPLESGQTVDPDGRTPDISHFVG-----GYGRSDMPTTISHVTNVNLLY-NR 630
Db 1373 STMEVKNTYKGAQVKTNGDSFVQDKNGVFTKDVSLNYGVNKKI--HVEPNDSGVYGN 1430
QY 631 QLNABEIRTLFLSODLIGTEAHMGSSGSSGSSRSTPGS 667
Db 1431 QGRITEL-TKDIEIDVIRQNNSSGGTGNNTSTSGS 1466

RESULT 15

NER1 MOUSE
ID NER1 MOUSE STANDARD; PRT; 409 AA.
AC C35657; OS5220; Q99KG5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Sialidase 1 precursor (EC 3.2.1.18) (Lysosomal sialidase) (N-acetyl-
DE alpha-neuraminidase 1) (G9 sialidase).
GN NEUL OR NEU.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/RIJ;
RX MEDLINE=98029867; PubMed=9363440;
RA Carrillo M.B., Milner C.M., Ball S.T., Snoek M., Campbell R.D.;
RT "Cloning and characterization of a sialidase from the murine
RT histocompatibility-2 complex: low levels of mRNA and a single amino
RT acid mutation are responsible for reduced sialidase activity in mice
RT carrying the Neula allele.";
RL Glycobiology 7:975-986(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=98046014; PubMed=9384611;
RA Igdoura S.A., Gafuik C., Mertinait C., Saberi F., Paheshetsky A.V.,
RA Potter M., Trasler J.M., Gravel R.A.;
RT "Cloning of the cDNA and gene encoding mouse lysosomal sialidase and
RT correction of sialidase deficiency in human sialidosis and mouse SM/J
RT fibroblasts.";
RL Hum. Mol. Genet. 7:115-121(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Rowen L., Madan A., Qin S., Hall J., Dahl T., James R., Dickhoff R.,
RA Schaffer T., Ratcliffe A., Abbasi N., Loretz C., Lasky S., Hood L.;
RT "Sequence of the mouse MHC class III region.";
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

RNA SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zesberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A.C., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RP VARIANT ILB-209.
RX MEDLINE=98087585; PubMed=9425240;
RA Rottier R.J., Bonten E.J., d'Azzo A.;
RT "A point mutation in the neu-1 locus causes the neuraminidase defect
RT in the SM/J mouse.";
RL Hum. Mol. Genet. 7:313-321(1998).
CC -!- FUNCTION: Catalyzes the removal of sialic acid (N-acetylneuramic
CC acid) moieties from glycoproteins and glycolipids. To be active, it
CC is strictly dependent on its presence in the multienzyme complex.
CC Appears to have a preference for alpha 2-3 and alpha 2-6 sialyl
CC linkage (By similarity).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-,
CC alpha-(2->8)-glycosidic linkages of terminal sialic residues in
CC oligosaccharides, glycoproteins, glycolipids, colominic acid and
CC synthetic substrates.
CC -!- SUBUNIT: Interacts with cathepsin A (protective protein), beta-
CC galactosidase and N-acetylglucosamine-6-sulfate sulfatase in a
CC multienzyme complex.
CC -!- SUBCELLULAR LOCATION: Lysosomal. Localized not only on the inner
CC side of the lysosomal membrane and in the lysosomal lumen, but
CC also on the plasma membrane and in intracellular vesicles.
CC -!- TISSUE SPECIFICITY: Highly expressed in kidney, epididymis,
CC followed by brain, spinal cord and weakly expressed in adrenal,
CC heart, liver, lung and spleen.
CC -!- DOMAIN: A C-terminal internalization signal (YGTL) appears to
CC allow the targeting of plasma membrane proteins to endosomes.
CC -!- PTM: N-glycosylated (Probable).
CC -!- PTM: Phosphorylation of tyrosine within the internalization signal
CC results in inhibition of sialidase internalization and blockage on
CC the plasma membrane.
CC -!- SIMILARITY: Belongs to family 33 of glycosyl hydrolases.
CC -!- SIMILARITY: Contains 4 BNR repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y11412; CAA72215.1; -;
DR EMBL; U93702; AAC53536.1; -;
DR EMBL; AF109906; AAC84167.1; -;
DR EMBL; BC004666; AA04666.1; -;
DR HSSP; Q02834; 1EUR.
DR MGD; MGI:97305; Neul.
DR InterPro; IPR002860; GH_BNR.
DR Pfam; PF02012; BNR; 4.

KW Hydrolase; Glycosidase; Signal; Repeat; Glycoprotein.

FT SIGNAL 1 41 POTENTIAL.

FT CHAIN 42 409 SIALIDASE 1.

FT REPEAT 106 117 BNR 1.

FT REPEAT 166 177 BNR 2.

FT REPEAT 225 236 BNR 3.

FT REPEAT 341 352 BNR 4.

FT SITE 71 74 FRIP MOTIF.

FT SITE 406 409 INTERNALIZATION SIGNAL.

FT ACT_SITE 72 72 By similarity.

FT ACT_SITE 91 91 Potential.

FT ACT_SITE 97 97 Potential.

FT ACT_SITE 129 129 Potential.

FT ACT_SITE 258 258 Potential.

FT ACT_SITE 274 274 Potential.

FT ACT_SITE 335 335 By similarity.

FT ACT_SITE 364 364 Potential.

FT ACT_SITE 388 388 Potential.

FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 337 337 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 346 346 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 372 372 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT VARIANT 209 209 L -> I (REDUCED ACTIVITY).

FT CONFLICT 107 107 R -> K (IN REF. 2).

FT CONFLICT 113 113 S -> C (IN REF. 2).

FT CONFLICT 117 121 STAF1 -> PTGFM (IN REF. 2).

FT CONFLICT 172 173 GI -> AF (IN REF. 2).

FT CONFLICT 344 344 P -> L (IN REF. 2).

FT CONFLICT 351 351 Q -> L (IN REF. 4).

SQ SEQUENCE 409 AA; 44591 MW; 416BFD5BE27B8893 CRC64;

Query Match 4.1%; Score 142.5; DB 1; Length 409;

Best Local Similarity 23.5%; Pred. No. 0.0082;

Matches 92; Conservative 50; Mismatches 153; Indels 97; Gaps 20;

QY 51 VHSFRPLALNVND-GYVVAIADARYETSNDNSLIDTVAKYSVDDGETWETQIAIKNSRAS 109

Db 68 VDTFRPLIATPRGTLAPAEARKKSASDEGAKFIAMRRSTDQGSTW-----S 116

QY 110 SVSRVVDPTVIVKGNKLYLVGNSRSYSWTSHGDARDWDILLAVGEVTKTAGGKITA 169

Db 117 STAFIVDDGEASDGLNLGAVVNDVTGIVFLI-----YTLCAHKVNCQVASTMLVW 167

QY 170 S-----IKWGSP-----VSLKEFFPAEMEGMHTNQFLGGAGVAIVASGNLVYPVQVTNK 219

Db 168 SKDDGISWSPPRNLSVDIGTEMPAGPGSGIQ-KQREPGKGRLLVCGHGT-----E 218

QY 220 KQVFSKIFYSEDEGKTKFEGRS-----DFGCSE--PVALEWEGKLIINTP- 265

Db 219 RDGVFCLL--SDDHGASHYGTGVSIGIPGQPKHDFNPDCQPYELP-DGSVIINARN 275

QY 266 ---VDYRRRLVYESSDMGNSWVEAVGTLGRVMGSPKSNQPGSQSFTT-----VTIEGMR 318

Db 276 QNNYHCKRIVLRSYD-----ACDTLR-----PRDVTDPDLVDPVVAAGALATSSG-- 322

QY 319 VMLFTHPLNFKRWLDRNLN---WLTDNQRIYNGQVIGDENSAYSSVLKYDD----- 370

Db 323 IVFFSNPAH-----PEFRVNLTLRMSFNSGTQKERVQVWPGSPGYSLSLTALNSTDCK 377

QY 371 ----KLYCLHEINSN---EVYSIVFARLVGEL 395

Db 378 KQPPQLFLVLEKGLNRYTESISMVKLSVYGTLL 409

Search completed: August 3, 2004, 09:39:00

Job time : 17 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 09:36:00 ; Search time 41 Seconds
(without alignments)
5148.334 Million cell updates/sec

Title: US-10-086-913-2

Perfect score: 3517

Sequence: 1 MGSSHHHHHSSGLVPRGSH.....EAHWGSSGSSRSTPGSGC 669

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rviro:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3360	95.5	1060	5 Q26964	Q26964 trypanosoma
2	3299	93.8	964	5 Q26963	Q26963 trypanosoma
3	3269	92.9	642	5 Q26966	Q26966 trypanosoma
4	3267	92.9	642	5 Q26968	Q26968 trypanosoma
5	3262	92.7	642	5 Q26969	Q26969 trypanosoma
6	3254	92.5	642	5 Q26967	Q26967 trypanosoma
7	3199	91.0	892	5 Q7YZX6	Q7YZX6 trypanosoma
8	3132.5	89.1	643	5 Q9BHJ5	Q9BHJ5 trypanosoma
9	3093.5	88.0	879	5 Q00773	Q00773 trypanosoma
10	2920	83.0	700	5 Q26965	Q26965 trypanosoma
11	2846	80.9	736	5 Q26850	Q26850 trypanosoma
12	2384.5	67.8	660	5 Q4049	Q4049 trypanosoma
13	2259	64.2	611	5 Q27063	Q27063 trypanosoma
14	2160.5	61.4	700	5 Q08672	Q08672 trypanosoma
15	1853	52.7	649	5 Q27064	Q27064 trypanosoma
16	1445	41.1	313	5 Q26887	Q26887 trypanosoma

17	1151	32.7	771	5 Q9GU83	Q9GU83 trypanosoma
18	1141	32.4	771	5 Q9GSF0	Q9GSF0 trypanosoma
19	1069	30.4	497	5 Q7YZT3	Q7YZT3 trypanosoma
20	1029.5	29.3	906	5 Q81672	Q81672 trypanosoma
21	976	27.8	742	5 Q86DL7	Q86DL7 trypanosoma
22	957	27.2	692	5 Q86DL6	Q86DL6 trypanosoma
23	938.5	26.7	706	5 P90605	P90605 trypanosoma
24	931	26.5	824	5 Q26851	Q26851 trypanosoma
25	925	26.3	723	5 Q08661	Q08661 trypanosoma
26	918.5	26.1	694	5 Q86DL8	Q86DL8 trypanosoma
27	918	26.1	835	5 Q26971	Q26971 trypanosoma
28	905.5	25.7	711	5 Q26853	Q26853 trypanosoma
29	903	25.7	1003	5 Q26889	Q26889 trypanosoma
30	902	25.6	719	5 Q27349	Q27349 trypanosoma
31	898	25.5	826	5 Q26852	Q26852 trypanosoma
32	893	25.4	1003	5 Q09060	Q09060 trypanosoma
33	881.5	25.1	960	5 Q05507	Q05507 trypanosoma
34	876.5	24.9	595	5 Q9U7F3	Q9U7F3 trypanosoma
35	858.5	24.4	702	5 Q7YUQ9	Q7YUQ9 trypanosoma
36	838	23.8	927	5 Q8T300	Q8T300 trypanosoma
37	838	23.8	1018	5 Q8T301	Q8T301 trypanosoma
38	837	23.8	852	5 Q7YZX9	Q7YZX9 trypanosoma
39	831	23.6	678	5 Q94798	Q94798 trypanosoma
40	829.5	23.6	953	5 Q96602	Q96602 trypanosoma
41	828.5	23.6	927	5 Q8T304	Q8T304 trypanosoma
42	807.5	23.0	726	8 Q8M369	Q8M369 trypanosoma
43	806.5	22.9	705	5 Q76141	Q76141 trypanosoma
44	792	22.5	786	5 Q77209	Q77209 trypanosoma
45	789	22.4	590	5 Q9GSV2	Q9GSV2 trypanosoma

ALIGNMENTS

RESULT 1

Q26964 PRELIMINARY; PRT: 1060 AA.

AC Q26964; (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Trans-sialidase.

GN TCS-154.

OS Trypanosoma cruzi.

OC Eukaryota; Eukaryota; Kinetoplastida; Trypanosomatidae; Trypanosoma.

OX NCBI_TaxID=5693;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Y;

RA Uemura H.;

RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.

RP [2]

RP SEQUENCE OF 178-378 FROM N.A.

RC STRAIN=Y;

RX MEDLINE=93010978; PubMed=1396577;

RA Uemura H., Schenkman S., Nussenzeig V., Eichinger D.;

RT "Only some members of a gene family in Trypanosoma cruzi encode proteins that express both trans-sialidase and neuraminidase activities."

RL EMBO J. 11:3837-3844(1992).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=Y;

RX MEDLINE=8844669; PubMed=8844669;

RA Smith L.E., Uemura H., Eichinger D.;

RT "Isolation and expression of an open reading frame encoding sialidase from Trypanosoma rangeli."

RL Mol. Biochem. Parasitol. 79:21-34(1996).

DR EMBL; D50685; BAA09334.1; -

DR PIR; S28409; S28409.

DR PDB; 1MS0; 25-MAR-03.

DR GO; GO:0004308; F:exo-alpha-sialidase activity; IEA.

DR GO; GO:0009405; P:pathogenesis; IEA.

Db 633 TANNVLLYNRLQNAEEIRTLFLSQDLIGTEAHMDSSSDSNHSTPST 679

RESULT 3

Q26966 PRELIMINARY; PRT; 642 AA.

AC Q26966;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Trans-sialidase.

OS Trypanosoma cruzi.

OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.

OX NCBI_TaxID=5693;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95021510; PubMed=7935611;

RA Campetella O.E., Uttaro A.D., Parodi A.J., Frasch A.C.;

RT "A recombinant trypanosoma cruzi trans-sialidase lacking the amino acid repeats retains the enzymatic activity.";

RL Mol. Biochem. Parasitol. 64:337-340(1994).

DR EMBL; L26499; AAA66352.1; --

DR PDB; 1MR5; 18-MAR-03.

DR GO; GO:0004308; F:exo-alpha-sialidase activity; IEA.

DR GO; GO:0009405; P:pathogenesis; IEA.

DR InterPro; IPR002860; GH_BNR.

DR InterPro; IPR008377; Sialidase_trypan.

DR Pfam; PF02012; BNR; 3.

DR PRINTS; PR01803; TCSIALIDASE.

SQ SEQUENCE 642 AA; 70593 MW; 65BD95DD7ADB222 CRC64;

Query Match 92.9%; Score 3269; DB 5; Length 642;

Best Local Similarity 97.8%; Pred. No. 2.1e-223;

Matches 626; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 21 MAPGSSRVELFKROSSKVPPEKGGKVTERRVHSPRLPALVNDGVMVAIADARYETSDN 80

Db 2 LAPGSSRVELFKROSSKVPPEKGGKVTERRVHSPRLPALVNDGVMVAIADARYETSDN 61

Qy 81 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIKGNKLVLVGSSYNSRSY 140

Db 62 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIKGNKLVLVGSSYNSRSY 121

Qy 141 TSHGDARDWDLILAAGEVTSTAGGKITASIKWGSVPVSLKEFPFPAEMEGMHTNQFLGAG 200

Db 122 TSHGDARDWDLILAAGEVTSTAGGKITASIKWGSVPVSLKEFPFPAEMEGMHTNQFLGAG 181

Qy 201 VAIVASGNLVYPVQVTKKKQVFSKIFYSEDEGKTWKFGGRSDFGCSRPVALEWEGKL 260

Db 182 VAIVASGNLVYPVQVTKKKQVFSKIFYSEDEGKTWKFGGRSDFGCSRPVALEWEGKL 241

Qy 261 IINTRVDYRRLVYESSDMGNSWVEAVGTLRSVWGSPKSNQPGSQSSFTAVTIEGRVM 320

Db 242 IINTRVDYRRLVYESSDMGNSWVEAVGTLRSVWGSPKSNQPGSQSSFTAVTIEGRVM 301

Qy 321 LFTHPLNFKGRWLDRNLMLTDNQRIYNGVQVSGIDENSAYSVLYKDKLYCLHEINS 380

Db 302 LFTHPLNFKGRWLDRNLMLTDNQRIYNGVQVSGIDENSAYSVLYKDKLYCLHEINS 361

Qy 381 NEVYSLVFAELVGLRLIIKSVLQSKWQNDSHLSICTPADPAASSSRGCGPATTVGLV 440

Db 422 NEVYSLVFAELVGLRLIIKSVLQSKWQNDSHLSICTPADPAASSSRGCGPATTVGLV 421

Qy 441 GFLSHSATKTEWEDAYRCVNASTANAEVRPNGLKFCAGVGGGALWPVSOQONORYEPANH 500

Db 482 GFLSHSATKTEWEDAYRCVNASTANAEVRPNGLKFCAGVGGGALWPVSOQONORYEPANH 481

Qy 561 MGKRYHVLTWANKIGSVYIDGEPLEGSGQTVPDTPDISHPFVGGYGRSDMPTISHV 620

Db 542 MGKRYHVLTWANKIGSEYIDGEPLEGSGQTVPDTPDISHPFVGGYGRSDMPTISHV 601

Qy 621 TANNVLLYNRLQNAEEIRTLFLSQDLIGTEAHMGSSSGSS 660

Db 602 TANNVLLYNRLQNAEEIRTLFLSQDLIGTEAHMDSSSDTS 641

RESULT 4

Q26968 PRELIMINARY; PRT; 642 AA.

AC Q26968;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Trans-sialidase.

OS Trypanosoma cruzi.

OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.

OX NCBI_TaxID=5693;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=Tulahuen;

RX MEDLINE=95354943; PubMed=7628705;

RA Cremona M.L., Sanchez D.O., Frasch A.C., Campetella O.;

RT "A single tyrosine differentiates active and inactive Trypanosoma cruzi trans-sialidases.";

RL Gene 160:123-128(1995).

DR EMBL; L38457; AAA99443.1; --

DR GO; GO:0004308; F:exo-alpha-sialidase activity; IEA.

DR GO; GO:0009405; P:pathogenesis; IEA.

DR InterPro; IPR008985; ConA_like_lec_gl.

DR InterPro; IPR002860; GH_BNR.

DR InterPro; IPR008377; Sialidase_trypan.

DR Pfam; PF02012; BNR; 3.

DR PRINTS; PR01803; TCSIALIDASE.

SQ SEQUENCE 642 AA; 70459 MW; 09F047336492081B CRC64;

Query Match 92.9%; Score 3267; DB 5; Length 642;

Best Local Similarity 97.7%; Pred. No. 2.9e-223;

Matches 625; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Qy 21 MAPGSSRVELFKROSSKVPPEKGGKVTERRVHSPRLPALVNDGVMVAIADARYETSDN 80

Db 2 LAPGSSRVELFKROSSKVPPEKGGKVTERRVHSPRLPALVNDGVMVAIADARYETSDN 61

Qy 81 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIKGNKLVLVGSSYNSRSY 140

Db 62 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIKGNKLVLVGSSYNSRSY 121

Qy 141 TSHGDARDWDLILAAGEVTSTAGGKITASIKWGSVPVSLKEFPFPAEMEGMHTNQFLGAG 200

Db 122 TSHGDARDWDLILAAGEVTSTAGGKITASIKWGSVPVSLKEFPFPAEMEGMHTNQFLGAG 181

Qy 201 VAIVASGNLVYPVQVTKKKQVFSKIFYSEDEGKTWKFGGRSDFGCSRPVALEWEGKL 260

Db 182 VAIVASGNLVYPVQVTKKKQVFSKIFYSEDEGKTWKFGGRSDFGCSRPVALEWEGKL 241

Qy 261 IINTRVDYRRLVYESSDMGNSWVEAVGTLRSVWGSPKSNQPGSQSSFTAVTIEGRVM 320

Db 242 IINTRVDYRRLVYESSDMGNSWVEAVGTLRSVWGSPKSNQPGSQSSFTAVTIEGRVM 301

Qy 321 LFTHPLNFKGRWLDRNLMLTDNQRIYNGVQVSGIDENSAYSVLYKDKLYCLHEINS 380

Db 302 LFTHPLNFKGRWLDRNLMLTDNQRIYNGVQVSGIDENSAYSVLYKDKLYCLHEINS 361

Qy 381 NEVYSLVFAELVGLRLIIKSVLQSKWQNDSHLSICTPADPAASSSRGCGPATTVGLV 440

Db 362 NEVYSLVFAELVGLRLIIKSVLQSKWQNDSHLSICTPADPAASSSRGCGPATTVGLV 421

Qy 441 GFLSHSATKTEWEDAYRCVNASTANAEVRPNGLKFCAGVGGGALWPVSOQONORYEPANH 500

Db 422 GFLSHSATKTEWEDAYRCVNASTANAEVRPNGLKFCAGVGGGALWPVSOQONORYEPANH 481

Qy 501 AFTLVASVTIHEVPSVASPLLGASLDSSGGKLLGLSYDEKHQWQPIYGSTFVPTPGSWE 560

```
Db 482 ALTLVASVTIHEVRAASPLLGASLDSSGGKLLGLSYDEKQWQIYGSTVPTPGSWE 541
QY 561 MGKRYHVVLTMANKIGSVYIDGEPLESGGQTVVPDGRTPDISHFYVGGYGRSDMPTISHV 620
Db 542 TGKRYHVVLTMANKIGSVYIDGEPLESGGQTVVPDGRTPDISHFYVGGYGRSDMPTISHV 601
QY 621 TVNNVLLYNQLNAEIRTLFLSQDLIGTEAHMGSSSSGS 660
Db 602 TVNNVLLYNQLNAEIRTLFLSQDLIGTEAHMGSSSDTS 641

RESULT 5
Q26969 ID Q26969 PRELIMINARY; PRT; 642 AA.
AC Q26969;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Trans-sialidase.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tulahuén;
RX MEDLINE=95354943; PubMed=7628705;
RA Cremona M.L., Sanchez D.O., Frasch A.C., Campetella O.;
RT "A single tyrosine differentiates active and inactive Trypanosoma
RT cruzi trans-sialidases.";
RL Gene 160:123-128(1995).
DR EMBL; L38463; AAA99444.1; -.
DR GO; GO:0004308; P:exo-alpha-sialidase activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR008985; ConA_like_1ec_g1.
DR InterPro; IPR002860; GH_BNR.
DR InterPro; IPR008377; Sialidase_trypan.
DR Pfam; PF02012; BNR; 3.
DR PRINTS; PR01803; TCSIALIDASE.
SQ SEQUENCE 642 AA; 70606 MW; F17728B24ECF96A4 CRC64;
```

```
Query Match 92.7%; Score 3262; DB 5; Length 642;
Best Local Similarity 97.5%; Pred. No. 6.5e-223;
Matches 624; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

QY 21 MAPGSSRVELPKROSSKVPFEGKGVTERVHSPRLPALVNVGVMVAIADARYETSDN 80
Db 2 LAPGSSRVELPKROSSKVPFEGKGVTERVHSPRLPALVNVGVMVAIADARYETSDN 61
QY 81 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIKGNKLYLVGSSYNSRSY 140
Db 62 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIKGNKLYLVGSSYNSRSY 121
QY 141 TSHGDARDWDLILAVGEVTKTAGGKITASIKWGPSVSLKEFPFAEMEGHNTNQFLGGAG 200
Db 122 TSHGDARDWDLILAVGEVTKTAGGKITASIKWGPSVSLKEFPFAEMEGHNTNQFLGGAG 181
QY 201 VAIASGNLVYPVQVTKKKQVFSKIFYSEDEGKTKWFGESDFGCSEPVVALEWEGKL 260
Db 182 VAIASGNLVYPVQVTKKKQVFSKIFYSEDEGKTKWFGESDFGCSEPVVALEWEGKL 241
QY 261 IINTRVDYRRRLVYESSDMGNSWEAVGTLRSVWGPSPKSNQPGSSSFATVIEGRVM 320
Db 242 IINTRVDYRRRLVYESSDMGNSWEAVGTLRSVWGPSPKSNQPGSSSFATVIEGRVM 301
QY 321 LFTHPLNFKGRLDRNLNLTQRIYNNQVQVSGIDENSAYSSVLYKDDKLYCLHEINS 380
Db 302 LFTHPLNFKGRLDRNLNLTQRIYNNQVQVSGIDENSAYSSVLYKDDKLYCLHEINS 361
QY 381 NEYVSLVFARLVGELRIKSVLQSWKNWDSHLSICTPADPAASERCGPAAVTGVLV 440
Db 362 NEYVSLVFARLVGELRIKSVLQSWKNWDSHLSICTPADPAASERCGPAAVTGVLV 421
```

```
QY 441 GFLSHSATKTWEDAYECVNASTANAEVRPNGLKFAVGCGGALWPVSQQQNORYHPANH 500
Db 422 GFLSHSATKTWEDAYECVNASTANAEVRPNGLKFAVGCGGALWPVSQQQNORYHPANH 481
QY 501 AFLTLVASVTIHEVPSVASPLLGASLDSSGGKLLGLSYDEKQWQIYGSTVPTPGSWE 560
Db 482 AFLTLVASVTIHEVPSVASPLLGASLDSSGGKLLGLSYDEKQWQIYGSTVPTPGSWE 541
QY 561 MGKRYHVVLTMANKIGSVYIDGEPLESGGQTVVPDGRTPDISHFYVGGYGRSDMPTISHV 620
Db 542 TGKRYHVVLTMANKIGSVYIDGEPLESGGQTVVPDGRTPDISHFYVGGYGRSDMPTISHV 601
QY 621 TVNNVLLYNQLNAEIRTLFLSQDLIGTEAHMGSSSSGS 660
Db 602 TVNNVLLYNQLNAEIRTLFLSQDLIGTEAHMGSSSDTS 641

RESULT 6
Q26967 ID Q26967 PRELIMINARY; PRT; 642 AA.
AC Q26967;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Trans-sialidase.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RA;
RX MEDLINE=95354943; PubMed=7628705;
RA Cremona M.L., Sanchez D.O., Frasch A.C., Campetella O.;
RT "A single tyrosine differentiates active and inactive Trypanosoma
RT cruzi trans-sialidases.";
RL Gene 160:123-128(1995).
DR EMBL; L38456; AAA99442.1; -.
DR GO; GO:0004308; P:exo-alpha-sialidase activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR008985; ConA_like_1ec_g1.
DR InterPro; IPR002860; GH_BNR.
DR InterPro; IPR008377; Sialidase_trypan.
DR Pfam; PF02012; BNR; 3.
DR PRINTS; PR01803; TCSIALIDASE.
SQ SEQUENCE 642 AA; 70453 MW; BC263FAAAF61EA6E CRC64;
```

```
Query Match 92.5%; Score 3254; DB 5; Length 642;
Best Local Similarity 97.3%; Pred. No. 2.4e-222;
Matches 623; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

QY 21 MAPGSSRVELPKROSSKVPFEGKGVTERVHSPRLPALVNVGVMVAIADARYETSDN 80
Db 2 LAPGSSRVELPKROSSKVPFEGKGVTERVHSPRLPALVNVGVMVAIADARYETSDN 61
QY 81 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIKGNKLYLVGSSYNSRSY 140
Db 62 SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIKGNKLYLVGSSYNSRSY 121
QY 141 TSHGDARDWDLILAVGEVTKTAGGKITASIKWGPSVSLKEFPFAEMEGHNTNQFLGGAG 200
Db 122 TSHGDARDWDLILAVGEVTKTAGGKITASIKWGPSVSLKEFPFAEMEGHNTNQFLGGAG 181
QY 201 VAIASGNLVYPVQVTKKKQVFSKIFYSEDEGKTKWFGESDFGCSEPVVALEWEGKL 260
Db 182 VAIASGNLVYPVQVTKKKQVFSKIFYSEDEGKTKWFGESDFGCSEPVVALEWEGKL 241
QY 261 IINTRVDYRRRLVYESSDMGNSWEAVGTLRSVWGPSPKSNQPGSSSFATVIEGRVM 320
Db 242 IINTRVDYRRRLVYESSDMGNSWEAVGTLRSVWGPSPKSNQPGSSSFATVIEGRVM 301
QY 321 LFTHPLNFKGRLDRNLNLTQRIYNNQVQVSGIDENSAYSSVLYKDDKLYCLHEINS 380
Db 302 LFTHPLNFKGRLDRNLNLTQRIYNNQVQVSGIDENSAYSSVLYKDDKLYCLHEINS 361
```



```
Db 302 MLFTHPLNFKRWLRDLNLWLTQNRINYNGVSGIDENSAYSVLYKDDKLYCLHEIN 361
Qy 380 SNEYVSLVFARLVGELRII KSVLQSWKNDWDLSSICTPADPAASSERGCSPAVTTVGL 439
Db 362 TDEYVSLVFARLVGELRII KSVLRSKWNDWDLSSICTPADPAASSERGCSPAVTTVGL 421
Qy 440 VGLSHSATKTEWEDAYRCVNASTANAEVPNGLKFAVGCGGALWPVSQGGQNRYPAN 499
Db 422 VGLSGNASQVWEDAYRCVNASTANAEVRNGLKFAVGCGGALWPVSQGGQNRYPAN 481
Qy 500 HATLVASVTIHEVPVSVASPLLGASLDSSGKKLLGLSLYDEKHQVPIYGSTPVTPTGSW 559
Db 482 HATLVASVTIHEAPVSVASPLLGASLDSSGKKLLGLSLYDEKHQVPIYGSTPVTPTGSW 541
Qy 560 EMGRYHVLTWANKIGSVYIDGELPLEGSGQTVVDPGRTPDISHFYVGGYGRSDMPTISH 619
Db 542 ETGRYHVLTWANKIGSVYIDGELPLEGSGQTVVDPGRTPDISHFYVGGYGRSDMPTISH 601
Qy 620 VTNNVLLNRLNABEIRTLFSLQDLIGTEAHMSSSSSS 660
Db 602 VTNNVLLNRLNABEIRTLFSLQDLIGTEAHMSSSDTS 642

RESULT 9
Q00773 ID Q00773 PRELIMINARY; PRT; 879 AA.
AC Q00773;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Shed-acute-phase-antigen.
GN SAPA.
OS Trypanosoma cruzi.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]_TaxID=5693;
RP SEQUENCE FROM N.A.
RC STRAIN=CAI;
RX MEDLINE=92049558; PubMed=1840626;
RA Polievick G.D., Affranchino J.L., Frasch A.C., Sanchez D.O.;
RT "The complete sequence of a shed acute-phase antigen of Trypanosoma cruzi.";
RL Mol. Biochem. Parasitol. 47:247-250 (1991).
DR EMBL; X57235;
DR F01; S23006;
DR GO; GO:0004308; F:exo-alpha-sialidase activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR008985; ConA like_1ec_g1.
DR InterPro; IPR002860; GH BNR.
DR InterPro; IPR008377; Sialidase-trypan.
DR Pfam; PF02012; BNR; 2.
DR PRINTS; PR01803; TCSIALIDASE.
SQ SEQUENCE 879 AA; 93939 MW; 5E7255CA4C94D66E CRC64;

Query Match 88.0%; Score 3093.5; DB 5; Length 879;
Best Local Similarity 91.5%; Pred. No. 9.7e-211;
Matches 592; Conservative 21; Mismatches 33; Indels 1; Gaps 1;

Qy 21 MAPGSRVLFKQSSKVPPEKGGKVTYRHSFRLPALVNDGVNVAIADARYETSDN 80
Db 33 LAPGSRVLFKQNSTVPEENGVRQVHSFRLPALVNDGVNVAIADARYETSDN 92
Qy 81 SLIDTVAKYSVDDGETWETQIAIKNSRASSVRVDPTVIVKGNLYLVGSSNSRSY 140
Db 93 SLIDTVKYSVDDGETWETQIAIKNSRASSVRVDPTVIVKGNLYLVGSSNSRSY 152
Qy 141 TSHGDARDWILLAVGEVTKSTAGGKITASIKWGSFVSLKEFFPAEMEGMHTNPLGCG 200
Db 153 TSHGDARDWILLAVGEVTKSVTDGKTYANIYWGFSVSLKEFFPAYMEGHTNPLGCG 212
Qy 201 VAIVASGNLVPVQVNTNKKQVFSKIFYSEDDGKTKWFGKGRSDFGCSEPVALEWEGKL 272
Db 261 IINTRVDYRRLVYESSDMGNSVVEAVGTLRSVRGSPKSNQPSQSPTAVTIEGRVM 320
Db 273 IINTRVDYRRLVYESSDTEKPMVEAVGTLRSRWGSPKSDQPSQSPTAVTIEGRVM 332
Qy 321 LFTHPLNFKRWLRDLNLWLTQNRINYNGVSGIDENSAYSVLYKDDKLYCLHEIN 380
Db 333 LFTHPLNFKRWLRDLNLWLTQNRINYNGVQLSIGDENSAYSVLYKDDKLYCLHEIN 392
Qy 381 NEVYSLVFARLVGELRII KSVLQSWKNDWDLSSICTPADPAASSERGCSPAVTTVGL 440
Db 393 NEVYSLVFARLVGELRII KSVLQSWKNDWDLSSICTPADPAASSERGCSPAVTTVGL 452
Qy 441 GFLSHSATKTEWEDAYRCVNASTANAEVPNGLKFAVGCGGALWPVSQGGQNRYPAN 500
Db 453 GFLSHSANKTKWADAYRCVDASTANAEVRNGLKFAVGCGGALWPVSQGGQNRYPAN 512
Qy 501 AFTLVASVTIHEVPVSVASPLLGASLDSSGKKLLGLSLYDEKHQVPIYGSTPVTPTGSWE 560
Db 513 AFTLVASVTIHEAPVSVASPLLGASLDSSGKKLLGLSLYDEKHQVPIYGSTPVTPTGSWE 572
Qy 561 MGKRYHVLTWANKIGSVYIDGELPLEGSGQTVVDPGRTPDISHFYVGGYGRSDMPTISH 620
Db 573 TGKRYHVLTWANKIGSVYIDGELPLEGSGQTVVDPGRTPDISHFYVGGYGRSDMPTISH 631
Qy 621 TVNNVLLNRLNABEIRTLFSLQDLIGTEAHMSSSSSSSRSTPGS 667
Db 632 TVNNVLLNRLNABEIRTLFSLQDLIGTEAHMSSSDSAHGTFST 678

RESULT 10
Q26965 ID Q26965 PRELIMINARY; PRT; 700 AA.
AC Q26965;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Trans-sialidase homolog (fragment).
GN TCTS-193.
OS Trypanosoma cruzi.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]_TaxID=5693;
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RX MEDLINE=97001677; PubMed=8844669;
RA Smith L.E., Uemura H., Eichinger D.;
RT "Isolation and expression of an open reading frame encoding sialidase from Trypanosoma rangeli.";
RL Mol. Biochem. Parasitol. 79:21-34 (1996).
DR EMBL; D50586; BAA09335.1; -.
DR GO; GO:0004308; F:exo-alpha-sialidase activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR002860; GH BNR.
DR InterPro; IPR008377; Sialidase-trypan.
DR Pfam; PF02012; BNR; 2.
DR PRINTS; PR01803; TCSIALIDASE.
DR NON TER 700
SQ SEQUENCE 700 AA; 77107 MW; B5E6CB448261796B CRC64;

Query Match 83.0%; Score 2920; DB 5; Length 700;
Best Local Similarity 86.7%; Pred. No. 1.4e-198;
Matches 556; Conservative 27; Mismatches 58; Indels 0; Gaps 0;

Qy 21 MAPGSRVLFKQSSKVPPEKGGKVTYRHSFRLPALVNDGVNVAIADARYETSDN 80
Db 26 LAPESRIELFKRKNSTVPEFDKGVTERVHSFRLPALVNDGVNVAIADARYETSDN 85
Qy 81 SLIDTVAKYSVDDGETWETQIAIKNSRASSVRVDPTVIVKGNLYLVGSSNSRSY 140
Db 86 SLIDTVKYSVDDGETWETQIAIKNSRASSVRVDPTVIVKGNLYLVGSSNSRSY 145
```

```
Db 213 VAIVASGNLVPVQVNTNKKQVFSKIFYSEDDGKTKWFGKGRSDFGCSEPVALEWEGKL 272
Qy 261 IINTRVDYRRLVYESSDMGNSVVEAVGTLRSVRGSPKSNQPSQSPTAVTIEGRVM 320
Db 273 IINTRVDYRRLVYESSDTEKPMVEAVGTLRSRWGSPKSDQPSQSPTAVTIEGRVM 332
Qy 321 LFTHPLNFKRWLRDLNLWLTQNRINYNGVSGIDENSAYSVLYKDDKLYCLHEIN 380
Db 333 LFTHPLNFKRWLRDLNLWLTQNRINYNGVQLSIGDENSAYSVLYKDDKLYCLHEIN 392
Qy 381 NEVYSLVFARLVGELRII KSVLQSWKNDWDLSSICTPADPAASSERGCSPAVTTVGL 440
Db 393 NEVYSLVFARLVGELRII KSVLQSWKNDWDLSSICTPADPAASSERGCSPAVTTVGL 452
Qy 441 GFLSHSATKTEWEDAYRCVNASTANAEVPNGLKFAVGCGGALWPVSQGGQNRYPAN 500
Db 453 GFLSHSANKTKWADAYRCVDASTANAEVRNGLKFAVGCGGALWPVSQGGQNRYPAN 512
Qy 501 AFTLVASVTIHEVPVSVASPLLGASLDSSGKKLLGLSLYDEKHQVPIYGSTPVTPTGSWE 560
Db 513 AFTLVASVTIHEAPVSVASPLLGASLDSSGKKLLGLSLYDEKHQVPIYGSTPVTPTGSWE 572
Qy 561 MGKRYHVLTWANKIGSVYIDGELPLEGSGQTVVDPGRTPDISHFYVGGYGRSDMPTISH 620
Db 573 TGKRYHVLTWANKIGSVYIDGELPLEGSGQTVVDPGRTPDISHFYVGGYGRSDMPTISH 631
Qy 621 TVNNVLLNRLNABEIRTLFSLQDLIGTEAHMSSSSSSSRSTPGS 667
Db 632 TVNNVLLNRLNABEIRTLFSLQDLIGTEAHMSSSDSAHGTFST 678
```


"Trypanosoma rangeli sialidase: cloning, expression and similarity to
T. cruzi trans-sialidase";
RL Glycobiology 7:1167-1173(1997).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 2,3-, 2,6- AND 2,8-GLYCOSIDIC
CC LINKAGES JOINING TERMINAL NON-REDUCING N- OR O-ACETYLNEURAMINYL
CC RESIDUES TO GALACTOSE, N-ACETYLHEXOSAMINE, OR N- OR O-ACYLATED
CC NEURAMINYL RESIDUES IN OLIGOSACCHARIDES, GLYCOPROTEINS,
CC GLYCOLIPIDS OR COLOMINIC ACID.
DR EMBL; U83180; AAC95493.1; -;
DR EMBL; U83183; AAC95495.1; -;
DR EMBL; U83181; AAC95494.1; -;
DR PDB; 1MZ5; 16-OCT-02.
DR PDB; 1MZ6; 16-OCT-02.
DR PDB; 1N1S; 07-JAN-03.
DR PDB; 1N1T; 07-JAN-03.
DR PDB; 1N1Y; 07-JAN-03.
DR GO; GO:0004308; F:exo-alpha-sialidase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR008985; ConA like_lec_gl.
DR InterPro; IPR002860; GH BNR.
DR InterPro; IPR008377; Sialidase-trypan.
DR Pfam; PF02012; ENR; 3.
DR PRINTS; PR01803; TCSIALIDASE.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 660 SIALIDASE.
FT SEQUENCE 660 AA; 71964 MW; B5D55765DC4423D0 CRC64;
Query Match 67.8%; Score 2384.5; DB 5; Length 660;
Best Local Similarity 69.9%; Pred. No. 1.4e-160;
Matches 445; Conservative 76; Mismatches 118; Indels 1; Gaps 1;
QY 19 SHMAPSSRVELFKRQSSKVPFEK-GGKVTIRVVHSLPALVNVDSVMAIADARYETS 77
DB 21 SALAPSSRVELFKRKNSTVPFEESNGTIRERVVHSPRIPTLVNVDSVMAIADARYETS 80
QY 78 NDSLIDTVAKYSVDDGETWETQIAIKNSRASSSVSRVVDPTVIKGNKLVLVGSYNSSR 137
DB 81 FDSFIETAVKYSVDDGATWNTQIAIKNSRASSSVSRVMDATVIKGNKLVLVGSFNKTR 140
QY 138 SYWTSHGARDWDILLAVGEVTKSTAGGKITASIKWGSVSLKEFPFAEMEGMHTNQFLG 197
DB 141 NSWTQHRDGDWEPDLLVVGVEVTKSAANGKTTATISWGPVSLKPLPAEPFDGLITKEFIG 200
QY 198 GAGVAIVASNGNLVYPVQVTKKKQVFSKIFYSDEGKTWKFGESDFCCSEPALEWE 257
DB 201 GVGAIVGSNGNLVYPQIADMGGRVFTKIMYSEDDGNTWKFAEGRSKFGCSPEVALEWE 260
QY 258 GKLIINTRDYRRRLVYESSDMGNSWVEAVGTLRSVWGPSPKSNQPGSQSSFTAVTIEGM 317
DB 261 GKLIINRVGNRLVYESSDMGKTWVEALGTLVSHVWNTSPISNQDCCSSFVAVTIEGK 320
QY 318 RVMLFTHPLNFKGRWLRDLNMLTQRIYNYGVQISIGDENSAYSVLYKDDKLYCLHE 377
DB 321 RVMLFTHPLNFKGRWLRDLNMLTQRIYNYGVQISIGDENSAYSVLYKDDKLYCLHE 380
QY 378 INSNEVSVLFARLVGELRIIKSVLOSQKWNDSHLSICTPADPAASSSSRGCGPAVTTV 437
DB 381 INTNDVSVLFRVFIQELQKMSVVRWKEEDNHLASICTPVVPATPPSKGCGAAVPTA 440
QY 438 GLVGFLSHSATKTEWEDAYRCVNASTANAEVNPGLKFAVGCGGALWPVSOQONQRYHF 497
DB 441 GLVGFLSHSANGSVWEDVYRCVDANVAERVENGLKFGVCGGAWPVARQGTIRYQF 500
QY 498 ANHAFILVASVTHVPSVASPLLGASLDSSGKKLLGLSYDEKHWQPIYGSTPVTPTG 557
DB 501 ANRFTLVATVTIDELPKGTSPLLGAGLEGPGDKLLGLSYDKNQRQWRPLYGAAPASPTG 560
QY 558 SWEMGRKHVLTWANKISVYIDGEPLESGGTVPVDPGRTPDISHYFYGGYGRSDMPTI 617
DB 561 SWELHKYHVLTWADRQSVVVDQPLAGSGNTVVRGATLPDISHYFIGGPRSKGAPT 620

QY 618 SHVTNVNVLNQLNAEEIRTLFLSQDLIGTEAHWGSSS 657
DB 621 SRVTNTVNLVNRRLNSSEIRTLFLSQDMIGTDGAGTAA 660
RESULT 13
Q27063 PRELIMINARY; PRT; 611 AA.
AC Q27063;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Sialidase (fragment).
OS Trypanosoma rangeli.
OC Eukaryota; Euzoenzoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
RX NCBI_TaxID=5698;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Venezuelan El Tocuyo;
RX MEDLINE=97001677; PubMed=8844669;
RA Smith L.E., Uemura H., Eichinger D.;
RT "Isolation and expression of an open reading frame encoding sialidase
from Trypanosoma rangeli.";
RL Mol. Biochem. Parasitol. 79:21-33(1996).
DR EMBL; U46072; AAC47213.1; -;
DR HSSP; P29768; 3SIL.
DR GO; GO:0004308; F:exo-alpha-sialidase activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR002860; GH BNR.
DR InterPro; IPR008377; Sialidase-trypan.
DR Pfam; PF02012; BNR; 3.
DR PRINTS; PR01803; TCSIALIDASE.
FT NON_TER 1 1
FT NON_TER 611 611
SQ SEQUENCE 611 AA; 66901 MW; 1EADB40F51AAC35A CRC64;
Query Match 64.2%; Score 2259; DB 5; Length 611;
Best Local Similarity 69.9%; Pred. No. 9.8e-152;
Matches 428; Conservative 68; Mismatches 114; Indels 2; Gaps 2;
QY 21 MAPGSSRVELFKRQSSKVPFEK-GGKVTIRVVHSLPALVNVDSVMAIADARYETSND 79
DB 1 LAPGSSRVELFKRKNSTVPFEESNGTIRERVVHSPRIPTLVNVDSVMAIADARYETSFD 60
QY 80 NSLIDTVAKYSVDDGETWETQIAIKNSRASSSVSRVVDPTVIKGNKLVLVGSYNSSRSY 139
DB 61 NSFIIETAVKYSVDDGATWNTQIAIKNSRASSSVSRVMDATVIKGNKLVLVGSFNKTRNS 120
QY 140 WTSHGARDWDILLAVGEVTKSTAGGKITASIKWGSVSLKEFPFAEMEGMHTNQFLGA 199
DB 121 WTQHRDGDWEPDLLVVGVEVTKSAANGKTTATISWGPVSLKPLPAEPFDGLITKEFVG 180
QY 200 GVAIVASNGNLVYPVQVTKKKQVFSKIFYSDEGKTWKFGESDFCCSEPALEWEKG 259
DB 181 GAAIVASNGNLVYPQIADMGGRVFTKIMYSEDDGNTWKFAEGRSKFGCSPEVALEWG 240
QY 260 LIINTRDYRRRLVYESSDMGNSWVEAVGTLRSVWGPSPKSNQPGSQSSFTAVTIEGHRV 319
DB 241 LIINRVGNRLVYESSDMGKTWVEALGTLVSHVWNTSPISNQDCCSSFVAVTIEGKRV 300
QY 320 MLFTHPLNFKGRWLRDLNMLTQRIYNYGVQISIGDENSAYSVLYKDDKLYCLHEIN 379
DB 301 MLFTHPLNFKGRWLRDLNMLTQRIYNYGVQISIGDENSAYSVLYKDDKLYCLHEIN 360
QY 380 SNEVSVLFRVFIQELQKMSVVRWKEEDNHLASICTPVVPATPPSKGCGAAVPTAGL 439
DB 361 TNDVSVLFRVFIQELQKMSVVRWKEEDNHLASICTPVVPATPPSKGCGAAVPTAGL 420
QY 440 VGFLSHSATKTEWEDAYRCVNASTANAEVNPGLKFAVGCGGALWPVSOQONQRYHPAN 499
DB 421 VGFLSHSANGSVWEDVYRCVDANVAERVENGLKFGVCGGAWPVARQGTIRYQFAN 480

```
Qy 500 HAFVLVASVTIHEVPSVASPLLGASLDSSGGKLLGLSYDEKHQMPIQYGSTPTVPTGWS 559
Db 481 YRETLVATVITIDPPKGMSPLLGAGLEAPHG-KLLGLSYDKNRQWRPLYGAAPSPGWS 539
Qy 560 EMGKRVHVLTMANKIGSVVYIDGEPLEGSGQTVVDPGRTPDISHFYVGGYGRSDMPITISH 619
Db 540 ELHKKHVLTLMADROQSVCCVDOPLAGSGNVTVRGATLPDISHFYIIGRPRSGAPTDSR 599
Qy 620 VTNNVLLNQR 631
Db 600 VTNVNLNRR 611

RESULT 14
Q08672 PRELIMINARY; PRT; 700 AA.
AC Q08672;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE LDG trans-sialidase homologue.
OS Trypanosoma rangeli.
OC Eukaryota; Eulenzooza; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5698;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LDS;
RX MEDLINE=94158961; PubMed=8114811;
RA Buschiazio A., Crenona M.L., Campetella O., Frasch A.C., Sanchez D.O.;
RT "Sequence of a Trypanosoma rangeli gene closely related to Trypanosoma
RL Mol. Biochem. Parasitol. 62:115-116 (1993).
DR HSP; P29768; 35IL.
DR GO; GO:0004308; F:exo-alpha-sialidase activity; IEA.
DR InterPro; IPR008377; Sialidase_trypan.
DR Pfam; PF02012; BNR; 3.
DR PRINTS; PR01803; TCSIALIDASE.
SQ SEQUENCE 700 AA; 77356 MW; BA34581AE138E97C CRC64;

Query Match 61.4%; Score 2160.5; DB 5; Length 700;
Best Local Similarity 64.5%; Pred. No. 1.2e-144;
Matches 417; Conservative 68; Mismatches 147; Indels 15; Gaps 4;

Qy 19 SHMAGSSRVLEPKROSSKVPPEK-GGKVTFRVHSPRLPALVNDGVMAIADARYETS 77
Db 23 SALAPGSSRVLEPKRONSTVPFENNNGTFRVRVDSFRIPALVNDGVMAVADARYETA 82

Qy 78 NDNSLIDTVAKYSVDDGETWETOIAIKNSRASSVSRVDPVTIVKGNKLYLVGSYNSSR 137
Db 83 NDNSFETVAKYSVDDGETWNTQIAIKNSRVSPSRVDPVTIVKGNKLYLVGSFNKSH 142

Qy 138 SYWTSHGARDWDILLAVGEVTKSTAGGKITASIKWGSVPVSLKEFPFAEMEGMHTNQFLG 197
Db 143 NYWQHDDGSDWEPLLVGGEVTKSTVNGKTTATISWGPVSLKAMPPEEGVPTKEYLG 202

Qy 198 GAGVAIVASNGNLYPVQVNTKKQVPSKIFPYSEDEGKTWKFGEGRSDFCCSEPALEWE 257
Db 203 GVGAIVASNGNLYPVQVNTIQRQITKIMYPEDDNGTWTWKFAGRSKFCSEPALEWE 262

Qy 258 GKLIINTRDYRRRLVYESSDMGNSVWEAVGTLSRVWGPSPKSNQPSQSSFTAVTIEGM 317
Db 263 GKLIINTRDYRRRLVYESSDMGNTWVEALGTLISRVWNTSPNSQPSQSSFFVAATIEGK 322

Qy 318 RVMLFTHPLNFKGRLDRNLNLTQRIYNYGOVSIIGDENSAYSVLYKDKLYCLHE 377
Db 323 RVMLFTHPLNFKGLWIRDLHLAVTDNQRIFDVQGISVSGEAGAAVSSVLYKDKLYCLHE 382

Qy 378 INSENVYSLVFAELVGLRIKSVLQSKWQNDHLSICTPADPAASSSRGCGPATTV 437
Db 383 VNREYVYIVFIRLVEELRILSVVGAWKEEDNLLASICTPVVPAAPSSRRGCGAAVPTA 442
```

```
Qy 438 GLVGFSLSHSATKTEWEDAYKVCNASTANAEVRPNGLKPAVGVGALWPVSQQQONQRYHF 497
Db 443 GLVGFSLSHRANKSVWEDVYRSVNATVMHGTNTVYGTFTKGRKAGAMWVRAKQGTDSAVLF 502
Qy 498 ANHAFVLVASVTIHEVPSVASPLLGASLDSSGGKLLGLSYDEKHQMPIQYGSTPTVPTG 557
Db 503 RKLQIHAGGLCIYHEIPNSTRPLLGASLDSSGNRKLGLSYDKDQRWCPWYGTPTASPTG 562
Qy 558 SWEMGKRYHVLTWANKIGSVYIDGEPLEGSGQTVVDPGRTPDISHFYVGGYGRSDM 614
Db 563 SWKLHQTTHVALLPDQGVGSIIYDGNLLKSGQILSGVHLEGL--DVSHEFFFRGYGTSDL 620
Qy 615 PTISHVTNNVLLNQRNLNABEIRTLFLSQDLIGTEAHMGSSSSGSE 661
Db 621 SSDCHITVTNNVLLNRRILKPNIEITLLS-----LGKTAADSE 658

RESULT 15
Q27064 PRELIMINARY; PRT; 649 AA.
AC Q27064;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Sialidase.
OS Trypanosoma rangeli.
OC Eukaryota; Eulenzooza; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5698;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Venezuelan El Tocuyo;
RX MEDLINE=97001677; PubMed=8844669;
RA Smith L.E., Uemura H., Eichinger D.;
RT "Isolation and expression of an open reading frame encoding sialidase
RL Mol. Biochem. Parasitol. 79:21-33 (1996).
DR EMBL; U46073; AAC47214.1; -.
DR GO; GO:0004308; F:exo-alpha-sialidase activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR002860; GH BNR.
DR InterPro; IPR008377; Sialidase_trypan.
DR Pfam; PF02012; BNR; 3.
DR PRINTS; PR01803; TCSIALIDASE.
SQ SEQUENCE 649 AA; 70872 MW; 26987D3786220004 CRC64;

Query Match 52.7%; Score 1853; DB 5; Length 649;
Best Local Similarity 61.3%; Pred. No. 7.2e-123;
Matches 374; Conservative 60; Mismatches 132; Indels 44; Gaps 8;

Qy 19 SHMAGSSRVLEPKROSSKVPPEK-GGKVTFRVHSPRLPALVNDGVMAIADARYETS 77
Db 53 SALAPGSSRVLEPKRONSTVPFENNNGTFRVRVDSFRIPALVNDGVMAVADARYETS 112

Qy 78 NDNSLIDTVAKYSVDDGETWETOIAIKNSRASSVSRVDPVTIVKGNKLYLVGSYNSSR 137
Db 113 FDNSTFETVAKYSVDDGETWNTQIAIKNSRASSVSRVMDATVIVKGNKLYLVGSFNKTR 172

Qy 138 SYWTSHGARDWDILLAVGEVTKSTAGGKITASIKWGSVPVSLKEFPFAEMEGMHTNQFLG 197
Db 173 NSWTHRDGSDWEPLLVGGEVTKSTVNGKTTATISWGPVSLKPLFLFAEPFDGILTKGFI 232

Qy 198 GAGVAIVASNG--NLVYPVQVNTKKQVPSKIFPYSEDEGKTWKFGEGRSDFCCSEPAVL 254
Db 233 GVG----GHRGVQWQVPCVPCANCGHGRVFTKIMYSEDDGNTWKFAGRSKFCSEPAVL 288

Qy 255 EWEKGLIINTRDYRRRLVYESSDMGNSVWEAVGTLSRVWGPSPKSNQPSQSSFTAVTI 314
Db 289 EWEKGLIINNVCGNRRLVYESSDMGNTWVEALGTLISRVWNTSPNSQPSQSSFFVAVTI 348

Qy 315 EGRVMLFTHPLNFKGRLDRNLNLTQRIYNYGOVSIIGDENSAYSVLYKDKLYC 374
Db 349 EGRVMLFTHPLNFKGRLDRNLNLTQRIYNYGOVSIIGDENSAYSVLYKDKLYC 408
```

```

Qy 375 LHEINSNEVYSLVFLVGLRIIISVLSQWKNWDSHLSSTCTPADPAASSSERGCGPAV 434
Db 409 LHEINTNDVYSLVFLVRLIGELQLMKSVVVTWKEGNHLASICTPVVLTATPPSKGAVG-AV 467
Qy 435 TTVGLVGLFSLHSATKTEWEDAYRCVNASTANAEVPMGLKFPAGYGGGALMPVSQQGQNR 494
Db 468 PTAGLVGLFSLHSANGSVWEDVYRCVDANVANAERVPMGLKLVGGGAVWVPRQGTTR 527
Qy 495 YHEANHAFTLVASVTIHEVPSVASPLLGASLD-----SSGGKKLLGLSYDEKHQWQPIYGS 550
Db 528 YQFANYRFTLVATVTIDEPPKGMSPLLGAGLERATPSSWGCRTTRIASG-----PLYGA 581
Qy 551 TP-----VTPTGSWEMGKRYHVVLTMANKIGSVYIDGEPLEGSGQTWVPDGR 598
Db 582 APASRQDHGNCRTSTTWCSRWRGTGALCTLMA-----SLCGSGNTVVRCATL 628
Qy 599 PDISHFYVGG 608
Db 629 PDISHFCIGG 638

```

Search completed: August 3, 2004, 09:39:53
Job time : 44 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 09:36:35 ; Search time 21 Seconds
(without alignments)

3064.387 Million cell updates/sec

Title: US-10-086-913-2

Perfect score: 3517

Sequence: 1 MGSSHHHHSSGLVPRGSH.....EAHMGSSGSSERSTPGSGC 669

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 2833366

Minimum DB seq length: 0

Minimum DB seq	length: 0
Maximum DB seq	length: 200000000

Post-processing: Minimum Match 0%

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:

```
2: pir2:*
```

3: dir3: *

4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3093.5	88.0	879	2	S23006	shed acute-phase a
2	2681	76.2	1162	2	JH0557	exo-alpha-sialidas
3	1057	30.1	200	2	S28409	transsialidase-neu
4	918	26.1	835	1	A45596	trypanastigote-spe
5	902	25.6	719	2	S44237	trypanastigote sur
6	893	25.4	1003	2	JH0823	FL-160-2 protein -
7	881.5	25.1	959	2	S32016	flagellum-associat
8	806.5	22.9	705	2	T30521	surface protein -
9	779.5	22.2	744	2	A45622	surface antigen gp
10	536.5	15.3	641	2	PH1919	FL-160-4 protein -
11	491.5	14.0	420	2	S25236	85K surface antigen
12	394.5	11.2	404	2	A37234	exo-alpha-sialidas
13	388	11.0	382	2	S01339	exo-alpha-sialidas
14	359	10.2	382	2	S32148	exo-alpha-sialidas
15	335	9.5	376	1	NMBEST	exo-alpha-sialidas
16	291.5	8.3	375	2	S11294	SAB5-1.3 protein -
17	235.5	6.7	345	2	A48458	gp85/sialidase hom
18	234.5	6.7	545	2	JC2500	exo-alpha-sialidas
19	226.5	6.4	1035	2	T30287	exo-alpha-sialidas
20	226.5	6.4	1035	2	G98063	exo-alpha-sialidas
21	217	6.2	233	2	S11292	SAB5-1.1 protein -
22	208	5.9	240	2	S11293	SAB5-1.2 protein -
23	205	5.8	694	2	I40866	exo-alpha-sialidas
24	200	5.7	93	2	A39378	exo-alpha-sialidas
25	195	5.5	1014	1	NMCLSS	exo-alpha-sialidas
26	184	5.2	93	2	B39378	exo-alpha-sialidas
27	171.5	4.9	647	2	A45244	exo-alpha-sialidas
28	165	4.7	98	2	S36035	FL-160 protein - T
29	162	4.6	648	2	T34999	probable neuramidase

ALIGNMENTS

RESULT 1

S23006

shed acute-phase antigen - Trypanosoma cruzi

C:Species: Trypanosoma cruzi

C;Date: 17-Apr-1993 #sequence revision 17-Apr-1993 #text change 26-Aug-1999

C;Accession: S23006

R;Pollevick, G.D.; Affranchino, J.L.; Frasch, A.C.C.; Sanchez, D.O.

Mol. Biochem. Parasitol. 47, 247-250, 1991

A;Title: The complete sequence of a shed acute-phase antigen of Trypanosoma cruzi

A;Reference number: S23006; MUID:92049558; PMID:1840626

A;Accession: S23006

A;Molecule type: DNA

A;Residues: 1-879 <POL>

A;Cross-references: EMBL:X57235; NID:g10943; PID:g10944

C;Superfamily: trypomastigote-specific surface antigen

Query Match 88.0%; Score 3093.5; DB 2; Length 879;

Best Local Similarity 91.5%; Pred. No. 2.5e-215;

Matches 592; Conservative 21; Mismatches 33; Indels 1; Gaps 1;

Qy	21	MAPGSSRVELFKROSSVPPEKGGCKVTERVVHSPRLPALVNDVGVMVAIDARYETSDND	80
Db	33	LAPGSSRVELFKRQNSVTPPEENGVEVQRUVHSPRLPALVNDVGVMVAIDARYETSDND	92
Qy	81	SLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIKGNKLYVLVGSYNSRSRYW	140
Db	93	SLIDTVVKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIKGNKIYVLVGSYNSSTSYW	152
Qy	141	TSHGDARDWILLAVGEVTKSTAGKKTASTIKWGSVPVSLKEFFPAEMSGMHTNQFLGGAG	200
Db	153	TSHGDARDWILLAVGEVTKSVKDCKTANTIWGSPVSLKEFFPAYMEGMHTNQFLGGAG	212
Qy	201	VAIIVASNGNLVYPVQVNTKKCKQVPSKIPYSEDEGKTWKFGRSDPGCSEPVALBWEGL	260
Db	213	VAIIVASNGNLVYPVQVNTKRGKQVPSKIPYSEDDGKTWKFGRSDPGCSEPVALBWEGL	272
Qy	261	IINTRVDYRRRLVYESSDMGNSWVEAVGTLISRVGSPSPKSNQPSQSQSFATVITIEGMRVW	320
Db	273	IINTRVDYRRRLVYESSDTEKWPVEAVGTLISRCWGPSPKSNQPSQSQSFATVITIEGMRVW	332
Qy	321	LFTHPLNFKGWLDRDLNMLTMDNQRINYNVGQVSIQDENSAYSSVLYKDDKLYCLHEINS	380
Db	333	LFTHPLNFKGWLDRDLNMLTMDNQRINYNVGQLSIGDENSAYSSVLYKDDKLYCLHEINS	392
Qy	381	NEVYSLVFARLVGBLRILIKSVLQSWKQWDSHLSISCTPADPAASSSRGCGPAAVTVGLV	440
Db	393	NEVYSLVFARLVGBLRILIKSVLQSWKQWDSHLSISCTPADPAASSSERVCGPAAVTVGLA	452
Qy	441	GFLSHSATKTEWEDAYRCVNASTANABRVPNGLKPAVGGGALWPVSOOGNQRYHPANH	500
Db	453	GFLSHSANKTKWDAYRCVDASTANABRVSNGLKPAGVGGALWPVSOOGNQRYRFPANH	512

QY 501 AFTLVASVTIHEVPSVASPLLGASLDSSGGKLLGLSYDEKQWQPIYGSTPVTPTGSWE 560
Db |||||
QY 513 AFTLVASVTIHEAPSVASPLLGASLDSSGGKLLGLSYDEKQWQPIYGSTPVTPTGSWE 572
Db |||||
QY 561 MGKRYHVLTWANKISVYIDGPIGLSGGTVPVDPDIPSHFYVGGYGRSDMPTISHV 620
Db |||||
QY 573 TGRKYHVLTWNRNGSVYDGLLKGSGQTVVPD-RTPDISHFYVGGYGRSDMPTISHV 631
Db |||||
QY 621 TVNNVLLYNRLNAEIRTLFLSQDLIGTEAHMGSSGSSERSTPGS 667
Db |||||
QY 632 TVNNVLLYNRLNAEIRTLFLSQDLIGTEAHMGSSDSSAHGTPST 678
Db |||||

RESULT 2

JH0557
exo-alpha-sialidase (EC 3.2.1.18) - Trypanosoma cruzi
N/Alternate names: neuraminidase
C/Species: Trypanosoma cruzi
C/Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 11-Jan-2000
C/Accession: JH0557
R;Pereira, M.E.A.; Mejia, J.S.; Ortega-Barria, E.; Matzilevich, D.; Prioli, R.P.
J. Exp. Med. 174, 179-191, 1991
A/Title: The Trypanosoma cruzi neuraminidase contains sequences similar to bacterial neu
A/Reference number: JH0557; MUID:91277609; PMID:1711561
C/Accession: JH0557
A/Molecule type: DNA
A/Residues: 1-1162 <PER>
A/Cross-references: GB:M61732; NID:g162302; PID:g162303
A/Note: The authors translated the codon TCT for residue 45 as Cys
C/Comment: This protein plays a role in parasite-host cell interaction.
C/Superfamily: trypomastigote-specific surface antigen
C/Keywords: glycoprotein; glycosidase; hydrolase
F:394/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 76.2%; Score 2681; DB 2; Length 1162;
Best Local Similarity 86.5%; Pred. No. 2.6e+185;
Matches 525; Conservative 22; Mismatches 50; Indels 10; Gaps 4;

QY 66 MVAIDARYETSNLSLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIYVGNK 125
Db |||||
QY 1 MVAIDARYETSNLSLIDTVAKYSVDDGETWETQIAIKNSRASSVSRVVDPTVIYVGNK 60
Db |||||
QY 126 LYVLVGSYNSSRSYWTSHGDARDWILLAVGEVTKSTAGGKITASIKWSPVSLKPFPPA 185
Db |||||
QY 61 LYVLVGSYNSSRSYWSHGDARDWILLAVGEVTKSTAGGKITASIKWSPVSLKPFPPA 120
Db |||||
QY 186 EMEGMHTNQFLGGAGVAIVASNGNLVYPQVTNKKIKQVFSKIFYSDDEGKTWKFGRSD 245
Db |||||
QY 121 EMEGMHTNQFLGGAGVAIVASNGNLVYPQVTNKKIKQVFSKIFYSDDEGKTWKFGRSD 180
Db |||||
QY 246 FGCSEPVALEWEGKLIINTRVDYRRRLVYESSDMGNSWAEVGTLSRVWGPSPKSNQPGS 305
Db |||||
QY 181 FGCSEPVALEWEGKLIINTRVDWKRRLVYESSDMKWPVEAVGTLSRVWGPSPKSNQPGS 240
Db |||||
QY 306 QSFTAVTIEGMRVMLTFPLNFKGRWLDRNLMLTDNQRINYNVGVQVSGIDENSAYS6V 365
Db |||||
QY 241 QTSFTAVTIEGMRVMLTFPLNFKGRVDRNLMLTDNQRINYNVGVQVSGIDENSAYS6V 300
Db |||||
QY 366 LYKDDKLYCLHEINSNEVSLPARLVGLRIIKSVLQSKWQNDLSHSICTPADPAASS 425
Db |||||
QY 301 LYKDDKLYCLHEINTDEVSLPARLVGLRIIKSVLRSWKW---TATCPAPAPLLIQ 356
Db |||||
QY 426 SERGCGPAVTVTG---LVGFLSHSATKTEWEDAVRCVNSTANAEVRPNGLKFGVCGG 481
Db |||||
QY 357 PLRRQRVVVPLSPRLVLAFCRQLPK-RMGGSFRVCNVNSTANAEVRNGLKFGVCGG 415
Db |||||
QY 482 ALWPVSQQQONQRYHPANHAFTLVASVTIHEVPSVASPLLGASLDSSGGKLLGLSYDEK 541
Db |||||
QY 416 ALWPVSQQQONQRYHPANHAFTLVASVTIHEAPRAASPLLGASLDSSGGKLLGLSYDEK 475
Db |||||
QY 542 HQWQPIYGTPTVPTGSEWGRYHVLTWANKISVYIDGPIGLSGGTVPVDPDIPD 601
Db |||||
QY 476 HQWQPIYGTPTVPTGSEWGRYHVLTWANKISVYIDGPIGLSGGTVPVDPDIPD 535
Db |||||

QY 602 SHFYVGGYGRSDMPTISHVTNNVLLYN-RQLNAEIRTLFLSQDLIGTEAHMGSSGSS 660
Db |||||
QY 536 SHFYVGGYKESDMPITISHVTNNVLLYNRLNAEIRTLFLSQDLIGTEAHMGSSDSS 595
Db |||||
QY 661 ERSTPGS 667
Db |||||
QY 596 AHSTPST 602
Db |||||

RESULT 3

S28409
transsialidase-neuraminidase - Trypanosoma cruzi
C/Species: Trypanosoma cruzi
C/Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 11-Jan-2000
C/Accession: S28409
R;Uemura, H.; Schenkman, S.; Nussenzweig, V.; Eichinger, D.
EMBO J. 11, 3837-3844, 1992
A/Title: Only some members of a gene family in Trypanosoma cruzi encode proteins that e;
A/Reference number: S28409; MUID:93010978; PMID:1396577
C/Accession: S28409
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-200 <DEM>
C/Superfamily: trypomastigote-specific surface antigen

Query Match 30.1%; Score 1057; DB 2; Length 200;
Best Local Similarity 100.0%; Pred. No. 4.3e+69;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 ITASIKWGPSVSLKEFFPAEMEGMHTNQFLGGAGVAIVASNGNLVYPQVTNKKQVFSK 226
Db |||||
QY 1 ITASIKWGPSVSLKEFFPAEMEGMHTNQFLGGAGVAIVASNGNLVYPQVTNKKQVFSK 60
Db |||||
QY 227 IFYSDDEGKTWKFGRSDFGCSEPVALEWEGKLIINTRVDYRRRLVYESSDMGNSWVEA 286
Db |||||
QY 61 IFYSDDEGKTWKFGRSDFGCSEPVALEWEGKLIINTRVDYRRRLVYESSDMGNSWVEA 120
Db |||||
QY 287 VGTLSRVWGPSPKSNQPGSQSSFTAVTIEGMRVMLTFPLNFKGRWLDRNLMLTDNQR 346
Db |||||
QY 121 VGTLSRVWGPSPKSNQPGSQSSFTAVTIEGMRVMLTFPLNFKGRWLDRNLMLTDNQR 180
Db |||||
QY 347 IYNVGVQVSGIDENSAYS6V 366
Db |||||
QY 181 IYNVGVQVSGIDENSAYS6V 200
Db |||||

RESULT 4

A45596
trypomastigote-specific surface antigen precursor - Trypanosoma cruzi
C/Species: Trypanosoma cruzi
C/Date: 22-Apr-1993 #sequence_revision 02-Jun-1994 #text_change 16-Jul-1999
C/Accession: A45596; S27857
R;Fouts, D.L.; Ruef, B.J.; Ridley, P.T.; Wrighteman, R.A.; Peterson, D.S.; Manning, J.E
Mol. Biochem. Parasitol. 46, 189-200, 1991
A/Title: Nucleotide sequence and transcription of a trypomastigote surface antigen gene
A/Reference number: A45596; MUID:92018016; PMID:1717846
C/Accession: A45596
A/Molecule type: DNA
A/Residues: 1-835 <POU>
A/Cross-references: EMBL:M58466; NID:g162314; PIDN:AAA30259.1; PID:g162315
A/Experimental source: strain Peru
A/Note: sequence extracted from NCBI backbone (NCBIN:61039, NCBIIP:61041)
C/Genetics:
A/Gene: TSA-1
C/Superfamily: trypomastigote-specific surface antigen
C/Keywords: surface antigen
F:1-28/Domain: signal sequence #status predicted <STG>
F:23-835/Product: trypomastigote-specific surface antigen #status predicted <MAT>

Query Match 26.1%; Score 918; DB 1; Length 835;
Best Local Similarity 35.6%; Pred. No. 4.4e+58;
Matches 250; Conservative 107; Mismatches 256; Indels 90; Gaps 28;

QY	89	YSV-----DGETWETQIAIKNSRASSVSRVVDPTVIVKGNKLVVLGVSYN	134
Db	87	QPLTTKTANTPEVLKKPKGETRVLBEGASEDQKKVD-VSRPTTAVKGSIDYMLAGAYS	145
QY	135	SSRSYWTSHG--DARDWDILLAGEVTKSTAGKITASIKWGSFVSLKEFPFPAEMEGMHT	192
Db	146	SKVATPGQVGASDATQLGLLLVKGSVNSDANN-----KIDW-----KYTESPPORLFGTOP	198
QY	193	N---QPLGGAGVAIVASNGNLVVPQVTKKKQVFSKIFYSEDEGKT-----W	237
Db	199	DSWTKLIGSGSGVKMDETLFPVVGSTK-----AVEGTEEDGKTVSLIHHSSDNTW	252
QY	238	KFGEGRSDFGCSFPALEWE-GKLIINTRVDVRRRLVYBSSDMNGSNWRAVGFTLSRVMPG	296
Db	253	KLKSGISDGCSDPSVVEWEDDKLMMMTACDGRRRRVYESGDKGESWTEALGTLRSVGN	312
QY	297	SPKSNQPGS-QSSFTAVTI-----EGMRVMLFTHPLNFKRWLRDLRLMLTQNRINYVG	351
Db	313	NKKGEEAKTVRSGFITATVGNDDGDKRNVMLVTLVPMPIITAREKGKPHLWLTQNTHNVDIG	372
QY	352	QVSIQDGENS--AYSSVLY-----KDDKLYCLHEINSNEYS--LVFARLVGBELRII	398
Db	373	PIS-GDDDKDVAASSLLYNSGENTNEKXEDELIALYEKDGKPKSPGMVSVLLTEQLKRV	431
QY	399	KSVLQSKWNWDSHLSICTP--ADPAASSSRGCGPAVTTVGLVGLFSLSHATKTWEEDAY	456
Db	432	KDVLTIWKEVDKRVSQLCTSLIAEKERASTDDVCSAVKITAGLVGLFSLGFSNFKTWKEY	491
QY	457	RCVNASTANAE---RVPNGLKFPAGVGGALWPVSQOQONRYHPANHAFTLVASVTIHEV	513
Db	492	LGVNATVNNKEGAVQVENGKFTGRGAGAEWPFSGQGENQLYHFANYFTLVATVSIHKM	551
QY	514	P--SVASPLLGASLSSGKK--LLGLSYDEKHQMPIY--GSTPTVPTGSGWEMGKRYHV	567
Db	552	PGSDTPIPLMGVRSNGDENNVLLGLSYDNEKKLVLCDDGKTEEL--STLIGTDKTYQV	610
QY	568	VLTMAINKI-GSVYIDGEPLEGSGQTVVPDQRTPDISHFYVGGY--RSDMPTISHVTNN	624
Db	611	AIMLRNGTQGSVVDQRVGNGBECELGESKEISHFYIGDGVNAANKBGVS-VTVTN	669
QY	625	VLLYNRQLNAEEI	637
Db	670	VLLYNRPDDETEI	682
RESULT 9			
surface antigen gp85 - Trypanosoma cruzi			
C:Species: Trypanosoma cruzi			
C>Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 26-Aug-1999			
C:Accession: A45622			
R:Takle, G.B.; Cross, G.A.			
Mol. Biochem. Parasitol. 48, 185-198, 1991			
A:Title: An 85-kilodalton surface antigen gene family of Trypanosoma cruzi encodes polyh			
A:Reference number: A45622; MUID:92107221; PMID:1762630			
A:Accession: A45622			
A>Status: preliminary			
A:Molecule type: mRNA			
A:Residues: 1-744 <TAK>			
A>Note: sequence extracted from NCBI backbone (NCBIN:74863, NCBIPI:74864)			
C:Superfamily: trypanostigote-specific surface antigen			
C:Keywords: surface antigen			
Query Match 22.2%; Score 779.5; DB 2; Length 744;			
Best Local Similarity 33.4%; Pred. No. 3.8e-48;			
Matches 230; Conservative 100; Mismatches 272; Indels 87; Gaps 28;			
QY	27	RVELFKRQSSKVPFEKGGKVTERRVHVSFRLPALVNVVDGVMAIADARYETSN-DNSL---	82
Db	80	RDFLFVPQKT-VLLPRGGNKKWDSPASPSLVSGGVIAAFAGHLSKKKDKKSTEPS	138
QY	83	IDTVAKYSYDDGETWETQIA-IXNS-RASSVSRVVD-----PTVIVKGNKLYVLIV	130

Db	139	SDAWAY-IDSAWENSTLVEGVKSTWQAHTVLGKVDGKERFDVVLPTTTTKDNKVFLLA	197
QY	131	GSYNSRSYWTSHGDARDWDIILLAGEVTKSTAGKITASIKWGSFVS-LKEFPFPAEMEG	189
Db	198	GSSVASNCVSWSHGGLK---LKLIVGDVRKPT-DSQSGRINWGEVQSPLNENSGAVQER	253
QY	190	MHTNOPLFGAGVAIVASNGNLVVPQVTKKKQVFSKIFYSEDEGKTWFKGEGRSDFGCS	249
Db	254	KLT-AFVASGGAGVLMEDGTIVFSLMARNEEDDVSMIIYSKDDGTWALSNSVSAKCV	312
QY	250	EPVALEWEKGLIINTRVDVRRRLVYBSSDMNGSNWAEAVGTLRSVWGPSKSNQPGSQSS-	308
Db	313	NPRIWEGSLMIWDCDEQK-VYVSRDMGTITWEAVGKLLGVW---VNSGSGASQDS	368
QY	309	-----FTAVTIEGRVMLFTHPLNFKRWLRDLN---LWLTQNRINYVGVQSI--GDE	358
Db	369	LHVDALITATIEGRVMLYQ---RGNISGENANPLYLWVTDN-RSPHVPVGMDNAEK	423
QY	359	NSAYSSVLYKDDKLYCLHEINSNEYSVLVFAIRLVGBELRIIKSVLQSKWNWDSHLSICTP	418
Db	424	EBLESALLYSDGKJHLQRRVSGEGSVISLSRLTEBELKEIESVLSTWAKDIPFSSLSIP	483
QY	419	ADPAASSSRGCGPAVTTVGLVGLFSLSHATKTWEEDAYECVNASTANAEVRPNGLKFAV	478
Db	484	-----TAGLVAVLSDAAGDGRWDEYLCL-ATVKNVAVKVDGDFQLTES	525
QY	479	GGGALWPVSQOQONRYHPANHAFTLVASVTIHEVPSVASPLLGASLSSGKKLIGLSY	538
Db	526	NSRVLSMNVNTRDNNLHVFLSHDFTVAIVIIQNVPSGKTSLLTATL-ANTESYTMGLSY	584
QY	539	DEKHQMPIY--GSTPTVPTGSGWEMGKRYHVVLTMANKIGSVYIDGEPLEGSGQTVVPDG	596
Db	585	TADNKWETIFKDKPPTTESRPWPKBYQVALMLQGGKASVYIDGRSL-GECEALLTDE	643
QY	597	RPDISHFHYVGGYGRSDMPTISH---VTNNVLLYNRQLNAEIRTLFELSQDLI-----	647
Db	644	KSLEFVHFCGACVQESSPTAAQTKVTVTNVLNRPJLST-EMRAI---KDRIPKRG	699
QY	648	-----GTE-AHMGSSSSGSSSRSTPGSG	668
Db	700	PGSQVEGGTERRHPIRIEGVRANAPVSGS	728
RESULT 10			
PH1919			
FL-160-4 protein - Trypanosoma cruzi (fragment)			
C:Species: Trypanosoma cruzi			
C>Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 01-Dec-2000			
C:Accession: PH1919; S32017			
R:Van Voorhis, W.C.; Barrett, L.; Koelling, R.; Farr, A.G.			
J. Exp. Med. 178, 681-694, 1993			
A:Title: FL-160 proteins of Trypanosoma cruzi are expressed from a multigene family and			
A:Reference number: JH0823; MUID:93340646; PMID:7688032			
A:Accession: PH1919			
A:Molecule type: DNA			
A:Residues: 1-641 <VAN>			
R:Van Voorhis, W.C.; Barrett, L.; Koelling, R.; Farr, A.G.			
submitted to the EMBL Data Library, February 1993			
A:Description: FL-160 proteins of Trypanosoma cruzi are expressed from a multigene fami			
A:Reference number: S32015			
A:Accession: S32017			
A>Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-289, 'L', 291-309, 'N', 311-641 <VA2>			
A:Cross-references: EMBL:X70950; NID:g11162; FID:g11163			
A:Keywords: glycoprotein			
F.115,130,224,575,590/Binding site: carbohydrate (Asn) (covalent) #status predicted			
Query Match 15.3%; Score 536.5; DB 2; Length 641;			
Best Local Similarity 35.6%; Pred. No. 1.1e-30;			
Matches 127; Conservative 56; Mismatches 133; Indels 41; Gaps 9;			

Query Match 11.0%; Score 388; DB 2; Length 382;
Best Local Similarity 30.4%; Pred. No. 2.7e-20;
Matches 109; Conservative 72; Mismatches 138; Indels 40; Gaps 16;

QY 54 FRLP--ALVNVGVMAIADARYETSDNSLIDTVAKYSVDGDETQTQIAKNRA--SS 110
Db |||:||||::||::||::||::||::||::||::||::||::||::||:
 36 FRPNIQLN-DGITLFDSIRYNPDHAYIDIASARSTDFGKTWSYNIAMKNNRIDST 94
QY |||:||||::||::||::||::||::||::||::||::||::||:
Db 111 VSRVVD-PTVIVGNKLVLVGVSNNRSRY-WTSHGDARDWDILLAVGEVTKTAGCKIT 168
 |||:||||::||::||::||::||::||::||::||::||::||:
Db 95 YSRAMDSTLTITNTVRILIIASSWNTNGNWAMTTTSRRSDWSVMQVIYSD-----D 144
QY 169 ASIKWGSPVSL-KEFFPAEMEGMHTNOFLGAGVAIVASNGNLVYPQVVT---NKKKQVP 224
Db 145 NGLTWSNKIDLTDSSRVKQNQPNTICWLGGVSGSIVMDGTIVMPAQISIRENNENNY 204
QY 225 SKIFYSEDEGKWKFGEGRDPFCSEPVALEWECKLIINTRDYR-RLLVYESDDMGNSW 283
Db 205 SLIIYSKDNGETWTGMNKVNPNSTNSNNVIIELDGALIMSTRYDYGVRAAIYSHDLGTTW 264
QY 284 VEAVGTLRSRMVGPSPKSNQPSOSSFP-TAVTIEGRVLMFTPLNFKGRLMRDLRLMLWLT 342
Db 265 -EIVEPLN--GKITLKSGSCQGCSFIKATTSNGHRIGLISAPKNTGHEYIRDIOIAYMI 320
QY 343 D-NORIYNVGOVSI-----GDE-NSAYSVVLYKDDKLYCLHEINSN-----EVSVLV 387
Db 321 DFDDLKSXGOBECIPYPEDGKLGCGGYSCLSFKNNHLGIVYEANGNIEYOQLTPYTSLI 379

RESULT 14
S32148
exo-alpha-sialidase (EC 3.2.1.18) - Clostridium perfringens
C:Species: Clostridium perfringens
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 19-May-2000
C:Accession: S32148
R:Chien, C.H.
submitted to the EMBL Data Library, February 1993
A:Description: Extracellular neuraminidase gene of Clostridium perfringens ATCC 10543:
A:Reference number: S32148
A:Accession: S32148
A:Molecule type: DNA
A:Residues: 1-382 <CHI>
A:Cross-references: EMBL:X71123; NID:g288070; PIDN:CAA50436.1; PID:g288071
C:Superfamily: trypanosigote-specific surface antigen
C:Keywords: glycosidase; hydrolase

Query Match 10.2%; Score 359; DB 2; Length 382;
Best Local Similarity 28.4%; Pred. No. 3.4e-18;
Matches 102; Conservative 70; Mismatches 147; Indels 40; Gaps 15;

QY 54 FRLP--ALVNVGVMAIADARYETSDNSLIDTVAKYSVDGDETQTQIAKNRA--SS 110
Db |||:||||::||::||::||::||::||::||::||::||::||:
 36 FRPNIQLN-DGITLFDSIRYNAPDHAYIDIASARSTDFGKTWSYNIAMKNNRIDST 94
QY |||:||||::||::||::||::||::||::||::||::||::||:
Db 111 VSRVVDPTVIVGNKLVLVGVSNNRSRYWTSHGDA--DWLDILLAVGEVTKTAGCKIT 168
 |||:||||::||::||::||::||::||::||::||::||::||:
Db 95 YSRAMDSTLTITNTVRILIIASSWNTNGNWAMTTTSARRSDWSVMQVIYSD-----D 144
QY 169 ASIKWGSPVSL-KEFFPAEMEGMHTNOFLGAGVAIVASNGNLVYPQVVT---NKKKQVP 224
Db 145 NGLTWSNKIDLTDSSRVKQNQPNTICWLGGVSGSITMDGTIVMPQSISARENNENNY 204
QY 225 SKIFYSEDEGKWKFGEGRDPFCSEPVALEWECKLIINTRDYR-RLLVYESDDMGNSW 283
Db 205 SLIIYSKDNGETWTGMNKVNPNSTNSNNVIIELDVALLIMSTRYDYGVRAAIYSHDLGTTW 264
QY 284 VEAVGTLRSRMVGPSPKSNQPSOSSFP-AVTIEGRVLMFTPLNFKGRLMRDLRLMLWLT 342
Db 265 -EIVEPLN--GKITLKSGSCQGCSFIIATTNSGKRIALISAPKNTGHEYIRDIOIAYMI 320
QY 343 D-NORIYNVGOVSI-----GDE-NSAYSVVLYKDDKLYCLHEINSN-----EVSVLV 387
Db 321 DFDDLKSXGOBECIPYPEDGKLGCGGYSCLSFKNNHLGIVYDFNGNIEYOQLPYTSLI 379

RESULT 15

NMBEST

exo-alpha-sialidase (EC 3.2.1.18) - Salmonella typhimurium

exon-alpha-sialidase (EC 3.2.1.18) - Salmonella typhimurium

C:Species: Salmonella typhimurium

C:Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text_change 05-Jan-1996

C:Accession: S20976; S23020; S23980; A33133

C:Royer, L.L.; Hamilton, A.C.; Steenbergen, S.M.; Vimr, E.R.

Mol. Microbiol. 6, 873-884, 1992

A:Title: Cloning, sequencing and distribution of the Salmonella typhimurium LT2 sialidase

A:Reference number: S20976; MUID:92292952; PMID:1602967

A:Accession: S20976

A:Molecule type: DNA

A:Residues: 1-376 <HOY>

A:Cross-references: EMBL:M55342

A:Accession: S23020

A:Molecule type: protein

A:Residues: 2-40 <HO2>

R:Warner, T.G.; Harris, R.; McDowell, R.; Vimr, E.R.

Biochem. J. 285, 957-964, 1992

A:Title: Photolabelling of Salmonella typhimurium LT2 sialidase. Identification of a peptide

A:Reference number: S23980; MUID:92359969; PMID:1295492

A:Accession: S23980

A:Status: preliminary

A:Molecule type: protein

A:Residues: 270-299,301-357 <WAR>

C:Genetics:

A:Gene: nanH

C:Superfamily: Salmonella typhimurium exo-alpha-sialidase

C:Keywords: Glycosidase; hydrolase

F:2-376/Product: exo-alpha-sialidase #status experimental <MAT>

Query Match 9.5%; Score 335; DB 1; Length 376;

Best Local Similarity 28.2%; Pred. No. 1.8e-16;

Matches 111; Conservative 60; Mismatches 151; Indels 72; Gaps 17;

Qy 30 LFRQSKVPFEKG-----GKTERVVISFRPALVVD-GVMVAIADARYTSDNLSID 84

Db 8 VFKAEGEHFTDQKNTIVGSGGGTKYFRIPACMTTSKGTIVVPADARHNTASQSFID 67

Qy 85 TVAKYGVDDGETWETQTAIKNSRASS-VSRVVDPTIVKVG---NKLVLVGSVNSRSY 139

Db 68 TAAARSTGGKTMNKKAIAINDRVNSKLSRVMDPTCIANIQGRETIILVMGVKNNDKT 127

Qy 140 WTSHGDA---RDWDLILAVGEVTKSTAGG-----KITASI-----KWGSPVSLKEFFPAEM 187

Db 128 WGAYRKADPTDMDLVL-----YKSTDGVTFSKVETNIHDTVKNGT----- 170

Qy 188 EGMHTNOFLGAGVAIVASGNLVYPQVTKKK--QVFSKIFSYSEDEGKTWKGEGRSD 245

Db 171 ----ISAMLGSGVGLQNDGKLVFPVQVMTKNTITVLTNSFIYSTDGTITWSLPSGYCE 226

Qy 246 FGCSEPVALEWEGKLIINTRVDYERRLVYESSDMGNSVVEAVGTLRSRVWGSPK---SNQ 302

Db 227 FGFSENNIIIFNASLVNIRNSGLRR-SFETKDFGKTWTE-----FPMDKKVDNRN 277

Qy 303 PGSOSSFTAVTIEGMRVMLFTHPLNFKGRMLRDLRLNLMLTDNQRIYNVQGVISGD----- 357

Db 278 HGVOGQSTITPSGNKLVAAHSSAQKNNDYTRSDISLYA---HNLYS-GEVKLIDAFYPK 333

Qy 358 ----ENSAYGSVLKGD--DKLYCLHEINSNEVYS 385

Db 334 VGNASGAGYSCLSYRKQVNDKKHCHMLSKMPEVLS 367

Search completed: August 3, 2004, 09:40:23
Job time : 23 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 8, 2004, 17:19:08 ; Search time 12408 Seconds
(without alignments)
7021.236 Million cell updates/sec

Title: US-10-086-913-1
Perfect score: 2010
Sequence: 1 atggcgagcagcatcatca.....cgccggatccggctgctaa 2010

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_uni:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_ini:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rod:*
- 36: em_htg_man:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1926.2	95.8	3183	3	D50685	D50685 Trypanosoma
2	1926.2	95.8	3183	6	AR261979	AR261979 Sequence
3	1900.6	94.6	2895	3	D50684	D50684 Trypanosoma
4	1886.6	93.9	1929	3	TRBTRSI	L26499 Trypanosoma
5	1886.6	93.9	1929	6	AR261980	AR261980 Sequence
6	1881.8	93.6	1929	3	TRBTRSI	L38457 Trypanosoma
7	1877	93.4	1929	3	TRBTRSI	L38456 Trypanosoma
8	1873.8	93.2	1929	3	TRBTRSI	AF525766 Trypanosoma
9	1809.4	90.0	30095	3	AF525766	AF525766 Trypanosoma
10	1782.4	88.7	1932	3	TCR276679	AJ276679 Trypanosoma
11	1756.6	87.4	2874	3	TCR276679	X57235 T. cruzi she
12	1753.8	87.3	2133	3	TCJ002174	AJ002174 Trypanosoma
13	1708.2	85.0	5403	3	TRBTRCNA	M61732 T. cruzi neu
14	1666	82.9	2100	3	D50686	D50686 Trypanosoma
15	1622.2	80.7	2749	3	TCU01098	U01098 Trypanosoma
16	1623.2	81.5	2620	3	TRU83180	U83180 Trypanosoma
17	1185.8	59.0	2277	3	TRU46074	U46074 Trypanosoma
18	1179	58.7	1833	3	TRU46072	U46072 Trypanosoma
19	1121.4	55.8	2203	3	TRU46073	U46073 Trypanosoma
20	1032.2	51.4	3245	3	TRBSIALTRN	L14943 Trypanosoma
21	847.4	42.2	940	3	TCU50162	U50162 Trypanosoma
22	601	29.9	601	3	TRB0B	D12741 Trypanosoma
23	585	29.1	601	3	TRB0A	D12740 Trypanosoma
24	339.6	16.9	1491	3	TC0535487	AJ535487 Trypanosoma
25	335	16.7	131541	2	AC113260	AC113260 Trypanosoma
26	327.8	16.3	2939	3	AY142111	AY142111 Trypanosoma
27	319	15.9	19801	2	AC104490	AC104490 Trypanosoma
28	314.6	15.7	42075	2	AC097056	AC097056 Trypanosoma
29	313	15.6	32554	2	AC096945	AC096945 Trypanosoma
30	307.8	15.3	32126	2	AC097058	AC097058 Trypanosoma
31	302.8	15.1	3300	3	AF092099	AF092099 Trypanosoma
32	302.8	15.1	21765	2	AC096778	AC096778 Trypanosoma
33	301.2	15.0	28208	2	AC097057	AC097057 Trypanosoma
34	292.4	14.5	2657	3	AF091620	AF091620 Trypanosoma
35	282.2	14.0	119452	2	AC104502	AC104502 Trypanosoma
36	276.6	13.8	2150	3	AF128843	AF128843 Trypanosoma
37	268.2	13.3	2897	3	AY186573	AY186573 Trypanosoma
38	264.2	13.1	3313	3	AF181287	AF181287 Trypanosoma
39	263.8	13.1	2079	3	AY186574	AY186574 Trypanosoma
40	260	12.9	2316	3	AF310232	AF310232 Trypanosoma
41	242.6	12.1	2232	3	TCU77951	U77951 Trypanosoma
42	240.4	12.0	2140	3	TRB82KDSA	L14824 Trypanosoma
43	240	11.9	3727	3	TRBTSAL	M58466 Trypanosoma
44	237.8	11.8	21700	3	AY298908	AY298908 Trypanosoma
45	235.6	11.7	2485	3	TC85SUPR	Z32744 T. cruzi gen

ALIGNMENTS

RESULT 1

D50685
LOCUS Trypanosoma cruzi 3183 bp DNA linear INV 10-JAN-2003
DEFINITION TCTS-154 gene for trans-sialidase, complete cds.
ACCESSION D50685
VERSION D50685.1 GI:840707
KEYWORDS Trypanosoma cruzi
SOURCE Trypanosoma cruzi
ORGANISM Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
REFERENCE 1 (bases 534 to 1134)
AUTHORS Uemura.H., Schenkman.S., Nussenzweig.V. and Eichinger.D.
TITLE Only some members of a gene family in Trypanosoma cruzi encode proteins that express both trans-sialidase and neuraminidase

JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	activities EMBO J. 11 (11), 3837-3844 (1992) 93010978 1396577 2 Smith, L.E., Uemura, H. and Eichinger, D. Isolation and expression of an open reading frame encoding sialidase from <i>Trypanosoma rangeli</i> Mol. Biochem. Parasitol. 79 (1), 21-33 (1996) 97001677 8844669 3 (bases 1 to 3183) Uemura, H. Direct Submission Submitted (27-MAY-1995) Haruki Uemura, Nagasaki University, Institute of Tropical Medicine, 1-12-4 Sakamoto, Nagasaki, 852, Japan (E-mail: H-Uemura@cc.nagasaki-u.ac.jp, Tel: 81-958-49-7837, Fax: 81-958-49-7805) Location/Qualifiers 1. .3183 /organism="Trypanosoma cruzi" /mol_type="genomic DNA" /strain="y" /db_xref="taxon:5693" /clone="154" /dev_stage="trypomastigote" 1. .3183 /gene="TCTS-154" 1. .3183 /gene="TCTS-154" /function="active trans-sialidase" /codon_start=1 /product="trans-sialidase" /protein_id="BAA09334.1" /db_xref="GI:840708" Translation="MGKVVGASRMFWLMEFVPLLLALCPSEPAHALAPGSRVLEFK RQSKPFEGGKQKTERVHSFRPLALVNDVGMVAIADARYETSDNLSLIDTVAKYS VDDGETWETQIALKNGRASVSVPDPTVIGKNKLYLVGVSYNRSRYSYWTGHDARD WDLLAVGETYKTAGKKTASIKWGPSVSLKEFFPAEMEGMHTNOFLGAGVAIVAS NGNLVYPQVQVTKKMGVFSKIFYSEDBEGKPKWEGKRSDFGCEPVLWEKLLINT RDVRRLLVSESDMGMSVVEAVGTLRSRVWGPSKPNQSGSSFTAVTIEGMRWLF THPLNFKRWLRDLNMLTDNORIYNGQVSIQDGNISVSVLTKDDKILCLHENS NEVYSIAPARLVGLERILIKSVLQSKNWDHLSICTPADPAASSESRGCGPAVTVH LVGFLSHASTBEDAYRCVNSTANAEVPLNGKLFAGVGGALPVPVQQQONRYH FANHAFTLVASVTHIRVPVAVSPLLGASLDSGGKLLGLSYDEKHQWPIYGSPTVY PTGSEWEMGRYHVLLMANKIGSVYIDGPLESGQTVVVDGRTPDISHFYVGGYGRS DMPTI SHVTNNVLLYNOLNABEIRTLFLSODLIGTRAHMGSSSSNAHSTPSPAD NGAHSPTSPADSSAHSPTSPADSSAHSPTSPADNGAHSPTSPGSSAHSPTSP ADNGAHSPTSPADSSAHSPTSPADNGAHSPTSPADNGAHSPTSPGSSAHSPTSP TPGDSNAHSPTSPADNGAHSPTSPADSSAHSPTSPADNGAHSPTSPADSSAHSPT PSTPADSSAHSPTSPADNGAHSPTSPADSSAHSPTSPADNGAHSPTSPADSSAHSPT STSPADSSAHSPTSPADSSAHSPTSPADSSAHSPTSPADNGAHSPTSPADSSAHSPT AHSPTSPADSSAHSPTSPADSSAHSPTSPADSSAHSPTSPADNGAHSPTSPADSS GAAFSAFSGGGLLCCAGALLHVFVMAVFP"
ORIGIN	
Query Match	95.8%; Score 1926.2; DB 3; Length 3183;
Best Local Similarity	99.6%; Pred. No. 0;
Matches 1931; Conservative	0; Mismatches 8; Indels 0; Gaps 0;
QY	62 TGGCACCCGGATCGAGCGAGTTGAGCTGTTTAAAGCGGCAAGCTCGAAGTGCCATTG 121
DB	98 TGGCACCCGGATCGAGCGAGTTGAGCTGTTTAAAGCGGCAAGCTCGAAGTGCCATTG 157
QY	122 AAAAGGCGGCAAAATCAACCGAGCGGTTGTCCACTCGTTCGCCCTCCCGCCCTTGTTA 181
DB	158 AAAAGGCGGCAAAATCAACCGAGCGGTTGTCCACTCGTTCGCCCTCCCGCCCTTGTTA 217
QY	182 ATGTGACGGGTGATGTTGTCATCGCGGAGCTCGCTACGAAACATCCAAATGACAACT 241
DB	218 ATGTGACGGGTGATGTTGTCATCGCGGAGCTCGCTACGAAACATCCAAATGACAACT 277
QY	242 CCCTCATTTGATACGGTGGCGAAGTACAGCTGGACGATGGGGAGACGTTGGGAGACCCAAA 301

DB	278 CCCTCATTTGATACGGTGGCGAAGTACAGCTGGACGATGGGAGACGTTGGGAGACCCAAA 337
QY	302 TTGCCATCAAGAACAGCTCGTGCATCTCTGTTTCTCTGTTGTTGGATCCACAGTGAATTG 361
DB	338 TTGCCATCAAGAACAGCTCGTGCATCTCTGTTTCTCTGTTGTTGGATCCACAGTGAATTG 397
QY	362 TGAAGGCGCAACAGCTTTTACGTCCTCGTTTGGAGAGCTTACAAAGTTTCGAGAGAGTACTGGA 421
DB	398 TGAAGGCGCAACAGCTTTTACGTCCTCGTTTGGAGAGCTTACAAAGTTTCGAGAGAGTACTGGA 457
QY	422 CGTCCGATGGTGTGATGCGAGAGACTGGGATATTCTGCTTCCCTTGGTGGAGTCAAGAGT 481
DB	458 CGTCCGATGGTGTGATGCGAGAGACTGGGATATTCTGCTTCCCTTGGTGGAGTCAAGAGT 517
QY	482 CCACTCGCGGCGCGCAAGATAAATCTGGAGTATCAAAATGGGGAGCCCGTGTCACTGAAGG 541
DB	518 CCACTCGCGGCGCGCAAGATAAATCTGGAGTATCAAAATGGGGAGCCCGTGTCACTGAAGG 577
QY	542 AATTTTCCCGCGGCAAAATGGAGGAATGCACACAAATCAATTTCTTGGCGGTGCAGGTG 601
DB	578 AATTTTCCCGCGGCAAAATGGAGGAATGCACACAAATCAATTTCTTGGCGGTGCAGGTG 637
QY	602 TTGCCATTTGTCGCTCCAAACCGGATCTTGTGTACCTGTGAGAGTTACGAAACAAAAGA 661
DB	638 TTGCCATTTGTCGCTCCAAACCGGATCTTGTGTGTACCTGTGAGAGTTACGAAACAAAAGA 697
QY	662 AGCAAGTTTTTTTCCAAAGATCTTCTACTCGGAAGAGAGGCGCAAGACGTGGAGTTTGGGG 721
DB	698 AGCAAGTTTTTTTCCAAAGATCTTCTACTCGGAAGAGAGGCGCAAGACGTGGAGTTTGGGG 757
QY	722 AGGTTAGAGTGAATTTTGGCTGCTCTGAACCTGTGGCCCTTGTAGTGGAGGGGAGGCTCA 781
DB	758 AGGTTAGAGTGAATTTTGGCTGCTCTGAACCTGTGGCCCTTGTAGTGGAGGGGAGGCTCA 817
QY	782 TCATAAACACATCGAGTTGACTATCGCCGCTCTGTTGTGTAGTGCAGTCCAGTGCATGGGA 841
DB	818 TCATAAACACATCGAGTTGACTATCGCCGCTCTGTTGTGTAGTGCAGTGCATGGGA 877
QY	842 ATTCTGGTGGAGGCTGTGCGCACGCTCTCACGCTGTGTGGGGCCCTCCACCAAAATCGA 901
DB	878 ATTCTGGTGGAGGCTGTGCGCACGCTCTCACGCTGTGTGGGGCCCTCCACCAAAATCGA 937
QY	902 ACCAGCCCGGAGTCAGAGAGCTTCTACTCGCTGACCATCGAGGGAATCGGTGTTATGC 961
DB	938 ACCAGCCCGGAGTCAGAGAGCTTCTACTCGCTGACCATCGAGGGAATCGGTGTTATGC 997
QY	962 TCTTCAACACACCCGCTGAAATTTTAAAGGGAAGTGGCTGCGCGACCGACTCAACTCTCTGGC 1021
DB	998 TCTTCAACACACCCGCTGAAATTTTAAAGGGAAGTGGCTGCGCGACCGACTCAACTCTCTGGC 1057
QY	1022 TGACGGATACCGAGCGATTATTAAGCTTGGGCAAGTATCCATTTGGTGTGATAAATTTCCG 1081
DB	1058 TGACGGATACCGAGCGATTATTAAGCTTGGGCAAGTATCCATTTGGTGTGATAAATTTCCG 1117
QY	1082 CCTACAGCTCCGCTCTGTACAAGGATGATAAGCTGTACTGTTTGGATGATGATCAACAGTA 1141
DB	1118 CCTACAGCTCCGCTCTGTACAAGGATGATAAGCTGTACTGTTTGGATGATGATCAACAGTA 1177
QY	1142 ACAGAGTGTACAGCTTGTGTTTGGCGGCTGTGTTGGCGAGCTACGGATCATTTAAATCAG 1201
DB	1178 ACAGAGTGTACAGCTTGTGTTTGGCGGCTGTGTTGGCGAGCTACGGATCATTTAAATCAG 1237
QY	1202 TGCTGAGTCTTGGGAAGATTTGGGACAGCACTGTCTCAGCATTTTGACACCCCTGCTGATC 1261
DB	1238 TGCTGAGTCTTGGGAAGATTTGGGACAGCACTGTCTCAGCATTTTGACACCCCTGCTGATC 1297
QY	1262 CAGCCGCTTTCGTCGTCAGAGCGTGGTTGTGTGCTCCGCTGTACACAGGTTGCTTTGTTG 1321
DB	1298 CAGCCGCTTTCGTCGTCAGAGCGTGGTTGTGTGCTCCGCTGTACACAGGTTGCTTTGTTG 1357
QY	1322 GCTTTTGTGTCGACAGTGCACCAAAAACGAATGGAGAGATGCGTACCGCTGCTGCAAG 1381

```
Db 1358 GCTTTTGTGCGACAGTGCCACCAAAACCGAATGGGAGGATGCGTACCGCTGGTCAACG 1417
Qy 1382 CAAGACGGCAAAATGCGAGAGGGTTCCGAACGGTTTGAAGTTTGGCGGGGTTGGCGGAG 1441
Db 1418 CAAGACGGCAAAATGCGAGAGGGTTCCGAACGGTTTGAAGTTTGGCGGGGTTGGCGGAG 1477
Qy 1442 GGGCGCTTTGGCGGTGAGCAGCAGGCGGAGAGATCAACGGTATCACTTTGCAAAACACG 1501
Db 1478 GGGCGCTTTGGCGGTGAGCAGCAGGCGGAGAGATCAACGGTATCACTTTGCAAAACACG 1537
Qy 1502 CGTTTACCGTGGTGGCGTCCGTTGACGATTCACGAGGTTCCGAGCGTCCGAGTCCCTTTCG 1561
Db 1538 CGTTTACCGTGGTGGCGTCCGTTGACGATTCACGAGGTTCCGAGCGTCCGAGTCCCTTTCG 1597
Qy 1562 TGGGTGCGAGCCCTGGAATCTTCTGTGGGCAAAAACCTCTGGGGCTCTCGTACGACGAGA 1621
Db 1598 TGGGTGCGAGCCCTGGAATCTTCTGTGGGCAAAAACCTCTCTGGGGCTCTCGTACGACGAGA 1657
Qy 1622 AGCACGATGGCAGCCATATACGGATCAACGCGCGGTGACGCCGACCGGATCGTGGGAGA 1681
Db 1658 AGCACGATGGCAGCCATATACGGATCAACGCGCGGTGACGCCGACCGGATCGTGGGAGA 1717
Qy 1682 TGGGTGAAGAGTACCACTGCTCTTACGATGCGAATAAAATTTGGTTCCGTTGTACATTG 1741
Db 1718 TGGGTGAAGAGTACCACTGCTCTTACGATGCGAATAAAATTTGGTTCCGTTGTACATTG 1777
Qy 1742 ATGAGAAACCTCTGGAGGGTTTCAGGGCAGACCGTTGTGCCAGACGGGAGACCGCTGACA 1801
Db 1778 ATGAGAAACCTCTGGAGGGTTTCAGGGCAGACCGTTGTGCCAGACGGGAGACCGCTGACA 1837
Qy 1802 TCTCCCACTTCTAGTTGGCGGTATGGAAGAGTGTATGCCAACATTAAGCCACGTGA 1861
Db 1838 TCTCCCACTTCTAGTTGGCGGTATGGAAGAGTGTATGCCAACATTAAGCCACGTGA 1897
Qy 1862 CGGTGAATAATGTTCTTCTTACACCGTCAAGTGCAGTGCAGGAGATCAGGACCTTGT 1921
Db 1898 CGGTGAATAATGTTCTTCTTACACCGTCAAGTGCAGTGCAGGAGATCAGGACCTTGT 1957
Qy 1922 TCTTGAGCCAGGACCTGATTGGCAGCGAAGCACACATGCGCAGCAGCGGAGCAGATG 1981
Db 1958 TCTTGAGCCAGGACCTGATTGGCAGCGAAGCACACATGCGCAGCAGCGGAGCAGATG 2017
Qy 1982 AAAGAAGTAGCCCGGATC 2000
Db 2018 CCCACAGTAGCCCTCAAC 2036

RESULT 2
AR261979
LOCUS AR261979
DEFINITION Sequence 1 from patent US 6323008.
ACCESSION AR261979
VERSION AR261979.1 GI:28073277
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 3183)
AUTHORS Pelletier, M., Barker, W.A., Hakes, D.J. and Zopf, D.A.
TITLE Methods for producing sialyloligosaccharides in a dairy source
JOURNAL Patent: US 6323008-A 1 27-NOV-2001;
FEATURES
    Location/Qualifiers
        source
            1..3183
            /organism="unknown"
            /mol_type="mRNA"
ORIGIN
Query Match 95.8%; Score 1926.2; DB 6; Length 3183;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1931; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Qy 62 TGGACCCGATCGAGCGGATGAGCTGTTTAAAGCGGCAAGCTCGAAGGTGCCATTG 121
|||||
```

```
Db 98 TGGCACCCGGATCGAGCCGAGTTGAGCTGTTTAAAGCGGCAAAAGCTCGAAGGTGCCATTG 157
Qy 122 AAAAGGGCGCAAAAGTCAACGAGCGGGTTGTCACATCGCTTTCGCGCTCCCGCCCTTGTGTA 181
Db 158 AAAAGGGCGGCAAAAGTCAACGAGCGGGTTGTCACATCGCTTTCGCGCTCCCGCCCTTGTGTA 217
Qy 182 ATCTGGACGGGGTGATGTTGCCATCGCGAGCGCTCGCTACGAAACATCAATGACAAC 241
Db 218 ATCTGGACGGGGTGATGTTGCCATCGCGAGCGCTCGCTACGAAACATCAATGACAAC 277
Qy 242 CCCTCATTTGATACGGTGGCGAAGTACAGCGTGGAGAGATGGGGAGACGTTGGGAGACCCAAA 301
Db 278 CCCTCATTTGATACGGTGGCGAAGTACAGCGTGGAGAGATGGGGAGACGTTGGGAGACCCAAA 337
Qy 302 TTGCCATCAAGAACAGTGTGTCATCGTCTGTTCTCGTGTGGTGGATCCCAAGTGTG 361
Db 338 TTGCCATCAAGAACAGTGTGTCATCGTCTGTTCTCGTGTGGTGGATCCCAAGTGTG 397
Qy 362 TGAAGGGCAACAGCTTTACGTCCTCGTTGGAGAGCTACACAGTTCCGAGGAGCTACTGGA 421
Db 398 TGAAGGGCAACAGCTTTACGTCCTCGTTGGAGAGCTACACAGTTCCGAGGAGCTACTGGA 457
Qy 422 CGTCGATGTTGATCGAGAGACTTGGGATATCTGCTTTCGCTTGGTGGAGTCAACGAAGT 481
Db 458 CGTCGATGTTGATCGAGAGACTTGGGATATCTGCTTTCGCTTGGTGGAGTCAACGAAGT 517
Qy 482 CCACTCGGGCGGCAAGATAAATCGAGATCAAAATGGGGAGCGCCGTTCTCACTGAAG 541
Db 518 CCACTCGGGCGGCAAGATAAATCGAGATCAAAATGGGGAGCGCCGTTCTCACTGAAG 577
Qy 542 AATTTTCCCGGGGAAATGGAAGGATGCACAAATCAATTTCTTGGCGGTGCGAGTG 601
Db 578 AATTTTCCCGGGGAAATGGAAGGATGCACAAATCAATTTCTTGGCGGTGCGAGTG 637
Qy 602 TTGCCATTTGGCGTCCAAACGGGAATCTTGTGTACCTGTGCAGGTTACGAAACAAAAGA 661
Db 638 TTGCCATTTGGCGTCCAAACGGGAATCTTGTGTACCTGTGCAGGTTACGAAACAAAAGA 697
Qy 662 AGCAAGTTTTTCCAAAGATCTTCTACTCGGAAGACGAGGGCAAGACGTTGGAAGTTGGG 721
Db 698 AGCAAGTTTTTCCAAAGATCTTCTACTCGGAAGACGAGGGCAAGACGTTGGAAGTTGGG 757
Qy 722 AGGTTAGGAGTGAATTTGGGCTGCTGCACTCTGGCCCTTGAGTGGGAGGGAAGCTCA 781
Db 758 AGGTTAGGAGTGAATTTGGGCTGCTGCACTCTGGCCCTTGAGTGGGAGGGAAGCTCA 817
Qy 782 TCATAAACACTCGAGTTGACTATCGCGCGCTCTGTTGTACGAGTCCAGTGAATGGGGA 841
Db 818 TCATAAACACTCGAGTTGACTATCGCGCGCTCTGTTGTACGAGTCCAGTGAATGGGGA 877
Qy 842 ATTCTGGGTGGAGGCTGTCGGCACGCTCTCACTGTGTGGGGCCCTCAACCAAAATCGA 901
Db 878 ATTCTGGGTGGAGGCTGTCGGCACGCTCTCACTGTGTGGGGCCCTCAACCAAAATCGA 937
Qy 902 ACCAGCCGCGAGTCAAGCAGCTTCACTCCCTGACCATCGAGGGAATCGGTGTTATGC 961
Db 938 ACCAGCCGCGAGTCAAGCAGCTTCACTCCCTGACCATCGAGGGAATCGGTGTTATGC 997
Qy 962 TCTTTCACACACCGCTGAAATTTTAAAGGAAGTGGCTGCGCGACCGGACTGAACCTCTGGC 1021
Db 998 TCTTTCACACACCGCTGAAATTTTAAAGGAAGTGGCTGCGCGACCGGACTGAACCTCTGGC 1057
Qy 1022 TGAAGGATAACACGCGCATTTATACGTTGGGCAAGTATCCATTTGGTGTGATAAAATTCG 1081
Db 1058 TGAAGGATAACACGCGCATTTATACGTTGGGCAAGTATCCATTTGGTGTGATAAAATTCG 1117
Qy 1082 CCTACAGTCCGTCCTGTACAGGATGATGATGCTGTTGTCATGAGATCAACAGTA 1141
Db 1118 CCTACAGTCCGTCCTGTGTACAGGATGATGATGCTGTTGTCATGAGATCAACAGTA 1177
Qy 1142 ACCAGGTGTACACCTTTGTTTTCGCGCTGTTGGCGAGCTACGATCATTTAAATCAG 1201
Db 1178 ACCAGGTGTACACCTTTGTTTTCGCGCTGTTGGCGAGCTACGATCATTTAAATCAG 1237
```


302 TTGCCATCAAGAAACAGTCGTGTCATCGTCTGTTCTCGTGTGTTGATGCCACAGTGATTG 361
Db |||||
338 TTGCCATCAAGAAACAGTCGTGTCATCGTCTGTTCTCGTGTGTTGATGCCACAGTGATTG 397
Qy |||||
362 TGAAGGGCAACAAGCTTTACGTCCTGTTGGAGCTTACAACAGTTTCAGAGAGCTACTGGA 421
Db |||||
398 TGAAGGGCAACAAGCTTTACGTCCTGTTGGAGCTTACAACAGTTTCAGAGAGCTACTGGA 457
Qy |||||
422 CGTGCATGCTGATCGAGAGACTGGGATATCTGTTGCGTTGTTGGTGGAGTCAACGAAGT 481
Db |||||
458 CGTGCATGCTGATCGAGAGACTGGGATATCTGTTGCGTTGTTGGTGGAGTCAACGAAGT 517
Qy |||||
482 CCACTCGGGCGGCAAGATAACTCGAGTATCAAAATGGGGAGCCCCGTTCTCACTGAAGG 541
Db |||||
518 CCACTCGGGCGGCAAGATAACTCGAGTATCAAAATGGGGAGCCCCGTTCTCACTGAAGG 577
Qy |||||
542 AATTTTCCGGCGGCAAGATAACTCGAGTATCAAAATGGGGAGCCCCGTTCTCACTGAAGG 601
Db |||||
578 AATTTTCCGGCGGCAAGATAACTCGAGTATCAAAATGGGGAGCCCCGTTCTCACTGAAGG 637
Qy |||||
602 TTGCCATGTTGGGCTCCAAACGGGAATCTGTTGACCTGTGACAGTTTACGAAACAAAAGA 661
Db |||||
638 TTGCCATGTTGGGCTCCAAACGGGAATCTGTTGACCTGTGACAGTTTACGAAACAAAAGA 697
Qy |||||
662 AGCAAGTTTTTTTCCAAAGATCTTCTACTCGGAAGACGAGGGCAAGAGTGTGGAAGTTTGGGG 721
Db |||||
698 AGCAAGTTTTTTTCCAAAGATCTTCTACTCGGAAGATGATGGAAGAGTGTGGAAGTTTGGGG 757
Qy |||||
722 AGGGTAGGAGTATTTGGTGTCTCTGAACCTGTGCCCCTTGAGTGAGGAGGGGAAGTCA 781
Db |||||
758 AGGGTAGGAGCGCTTTTGGCTGTCTGAAGCTGTGCCCCTTGAGTGAGGAGGGGAAGTCA 817
Qy |||||
782 TCATAAACACTCAGTTGACTATCGCCCGCTGTGTTGACAGTCCAGTGACATGGGA 841
Db |||||
818 TCATAAACACTCAGTTGACTATCGCCCGCTGTGTTGACAGTCCAGTGACATGGGA 877
Qy |||||
842 ATTCTGGGGTGAAGGCTGTGCGGACGCTCTCAAGTGTGTTGGGCCCCCTCACCAAAATCGA 901
Db |||||
878 ATACGTGGCTGGAAGGCTGTGCGGACGCTCTCAAGTGTGTTGGGCCCCCTCACCAAAATCGA 937
Qy |||||
902 ACCAGCCGGCAGTCAGAGAGCTTCACTGCCGTGACCATCGAGGGAATCGTGTGATGC 961
Db |||||
938 ACCAGCCGGCAGTCAGAGAGCTTCACTGCCGTGACCATCGAGGGAATCGTGTGATGC 997
Qy |||||
962 TCTTTCACACACCCGCTGAATTTTAAAGGAGGTGGCTGCGGACCGACTGAACCTCTGCG 1021
Db |||||
998 TCTTTCACACACCCGCTGAATTTTAAAGGAGGTGGCTGCGGACCGACTGAACCTCTGCG 1057
Qy |||||
1022 TGACGGATAACCAAGCGCATTTATAACGTTGGGCAAGTATCCATTTGGTGTGATGAAAATTCGG 1081
Db |||||
1058 TGACGGATAACCAAGCGCATTTATAACGTTGGGCAAGTATCCATTTGGTGTGATGAAAATTCGG 1117
Qy |||||
1082 CTTACAGCTCCGCTCTGTACAAGATGATGAAGCTGTACTGTTTGCATGAGATCAACAGTA 1141
Db |||||
1118 CCCACAGCTCCGCTCTGTACAAGATGATGAAGCTGTACTGTTTGCATGAGATCAACAGTA 1177
Qy |||||
1142 ACCAGGTACAGCTTGTGTTTGGCGCTGTTGGCGAGCTACGGATCAATTAATCAG 1201
Db |||||
1178 ACCAGGTACAGCTTGTGTTTGGCGCTGTTGGCGAGCTACGGATCAATTAATCAG 1237
Qy |||||
1202 TGTGTGAGTCTCTGGAAGAAATTTGGACAGCCACCTGTCCAGCATTTGACCCCTGCTGATC 1261
Db |||||
1238 TGTGTGAGTCTCTGGAAGAAATTTGGACAGCCACCTGTCCAGCATTTGACCCCTGCTGATC 1297
Qy |||||
1262 CAGCCGCTTCTGTCAGAGCGTGTGTTGGTCCCGCTGTCAACAGGTTGGTCTTTGTTG 1321
Db |||||
1298 CAGCCGCTTCTGTCAGAGCGTGTGTTGGTCCCGCTGTCAACAGGTTGGTCTTTGTTG 1357
Qy |||||
1322 GCTTTTGTGCGACAGTGCCACAAAACCGAATGGGAGGATGGTACCGTGTGCTCAACG 1381
Db |||||
1358 GCTTTTGTGCGACAGTGCCACAAAACCGAATGGGAGGATGGTACCGTGTGCTCAACG 1417

Qy 1382 CAAGCAACGCAATGCGAGAGGGTTCCGAACGGTTTGAAGTTTGGGGGGTTGGCGGAG 1441
Db |||||
1418 CAAGCAACGCAATGCGAGAGGGTTCCGAACGGTTTGAAGTTTGGGGGGTTGGCGGAG 1477
Qy 1442 GGGCGCTTTGGCGGTTGAGCCAGCAGGGGAGAGATCAACGGTATCACTTTGCAAAACCAAG 1501
Db |||||
1478 GGGCGCTTTGGCGGTTGAGCCAGCAGGGGAGAGATCAACGGTATCACTTTGCAAAACCAAG 1537
Qy 1502 CGTTCAACGCTGTTGGCGGTTGAGCCAGCAGGGTTCGAGCGTTCGAGCGTTCGCGAGTCTTTTC 1561
Db |||||
1538 CGTTCAACGCTGTTGGCGGTTGAGCCAGCAGGGTTCGAGCGTTCGAGCGTTCGCGAGTCTTTTC 1597
Qy 1562 TGGGTGCGAGCTTGGAGCTCTTCTGTTGGGCAAAAACCTCTGGGGCTCTCTGAGACGAGA 1621
Db |||||
1598 TGGGTGCGAGCTTGGAGCTCTTCTGTTGGGCAAAAACCTCTGGGGCTCTCTGAGACGAGA 1657
Qy 1622 AGCACAGTGGCAGCAATATACGGATCAACGCGGTGACGCCGACCGGATCGTGGGAGA 1681
Db |||||
1658 AGCACAGTGGCAGCAATATACGGATCAACGCGGTGACGCCGACCGGATCGTGGGAGA 1717
Qy 1682 TGGGTGAAGAGTACCAACGCTGTTCTTACGATGGCGGAATAAAAATTTGGTTCGGTGTACATTG 1741
Db |||||
1718 CGGTTAAGAGTACCAACGCTGTTCTTACGATGGCGGAATAAAAATTTGGTTCGGTGTACATTG 1777
Qy 1742 ATGAGAACTCTGAGGGTTTCAAGGCAGACCTTTGTGCAGACGGGAGGACGCTGACA 1801
Db |||||
1778 ATGAGAACTCTGAGGGTTTCAAGGCAGACCTTTGTGCAGACGGGAGGACGCTGACA 1837
Qy 1802 TCTCCCACTTACGTTGGGGGTATGGAAGGATGATATGCCAACCAATTAAGCCAGTGA 1861
Db |||||
1838 TCTCCCACTTACGTTGGGGGTATGGAAGGATGATATGCCAACCAATTAAGCCAGTGA 1897
Qy 1862 CGGTGAATTAATGTTCTTTTACAACCGTCAGCTGAATGCCAGGAGATCAGGACCTTTGT 1921
Db |||||
1898 CGGTGAATTAATGTTCTTTTACAACCGTCAGCTGAATGCCAGGAGATCAGGACCTTTGT 1957
Qy 1922 TCTTGAGCAGGACCTTGATTTGGCAGCAACATATGGGACAGCAGCGGAGAGTG 1981
Db |||||
1958 TCTTGAGCAGGACCTTGATTTGGCAGCAACATATGGGACAGCAGCAGCAGCAATG 2017
Qy 1982 AAGAAGTAGCCCGGATC 2000
Db 2018 CCCACAGTAGCCCTCAAC 2036

RESULT 4
TRBTRSI
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

TRBTRSI Trypanosoma cruzi trans-sialidase gene, complete cds.
L26499
L26499.1 GI:432484
trans-sialidase.
Trypanosoma cruzi
Trypanosoma cruzi
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
1 (bases 1 to 1929)
Campetella, O.E., Uttaro, A.D., Parodi, A.J. and Frasch, A.C.
A recombinant Trypanosoma cruzi trans-sialidase lacking the amino
acid repeats retains the enzymatic activity
Mol. Biochem. Parasitol. 64 (2), 337-340 (1994)
95021510
7935611
COMMENT Original source text: Trypanosoma cruzi trypanastigote DNA.
FEATURES
Location/Qualifiers
1..1929
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/db_xref="taxon:5693"
/dev_stage="trypanastigote"
1..1929
/codon_start=1
/product="trans-sialidase"

CDS

```
/protein_id="AAA66352.1"
/db_xref="GI:432485"
/translation="MLAPGSSRVELFKROSSKVPPEKQKQKVTERRVHSLPALNVND
GVMAVADARYETSDNSLIDTVAKSYDDGETWETQIALKNSRASSVSRVDDPTIV
KGNKLVVLGVSYNRSYNTSHGDARDWDILAVGEVTKSTAGGKITASIKWGSPLV
KEFPPEMEGHNTNFTLGAGVAIVASNGNLVYPQVNTKKQVFSKIFYSEDEGKTW
KFGKSAFGCSFPALEWEGKLIINTRDYRRRLVYSSDMGNSWLBVGLTSLRVWG
PSPKNOPQSSFTAVTIEGRVLMFTPLNPKRWLRDLNLWLTONTORLYNVGOV
SIDENSAYSVLYKDDKLYCLHEINSEVYSLVPAFLVGLR11KSVLQSWKWDHSH
LSSICTPADPAASSBRGCPAVTVGLVGLFSLHNTKTEREDAYCNVASTANHERV
PNGLKEPAGVGGALWPSOQGNQRYEPANHAFTVAVSVTHIEVPSVASPLIGASLOS
SGSKLLGLHSIDERHQQPIYGSPTPTGSMWEGKRVHVLVWANKIKGSYIDIEPL
EGSGQTVPEDEPTDISHFYGYIKRSMDPTISHVTNNVLLYNQLNAEBIRTLFLS
QDLIGTEAHWSSSDTSA"
```

ORIGIN

```
Query Match      93.9%; Score 1886.6; DB 3; Length 1929;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1901; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 62 TGGCACCCGGATCGAGCCGAGTTGAGCTGTTTAAAGCGCAAGCTCGAAGGTGCGATTG 121
DB 5 TGGCACCCGGATCGAGCCGAGTTGAGCTGTTTAAAGCGCAAGCTCGAAGGTGCGATTG 64

QY 122 AAAAGGGCGCAAGCTCACCGAGCGGGTGTCCACTCGTTCGGCTCCCGCCCTTGTTA 181
DB 65 AAAAGGACGGCAAGCTCACCGAGCGGGTGTCCACTCGTTCGGCTCCCGCCCTTGTTA 124

QY 182 ATGTGGAGGGGTGATGTTGTCATCGCGGACGCTCGCTACGAAACATCCAAAGCAACT 241
DB 125 ATGTGGACGGGTGATGTTGCCATCGCGGACGCTCGCTACGAAACATCCAAAGCAACT 184

QY 242 CCTCATTTGATACGTTGCGGAAGTACAGCTGGAAGTGAAGTGGGAGACGCCAAA 301
DB 185 CCTCATTTGATACGTTGCGGAAGTACAGCTGGAAGTGAAGTGGGAGACGCCAAA 244

QY 302 TTGCCATCAAGAACAGTCGTGCATCGTCTGTTTCTCGTGTGGTGGATGCCACAGTGTG 361
DB 245 TTGCCATCAAGAACAGTCGTGCATCGTCTGTTTCTCGTGTGGTGGATGCCACAGTGTG 304

QY 362 TGAAGGGCAACAGCTTTACGTCGTGTGGAAGTCAACAGTTCGAGGAGTACTGGA 421
DB 305 TGAAGGGCAACAGCTTTACGTCGTGTGGAAGTCAACAGTTCGAGGAGTACTGGA 364

QY 422 CGTCGCATGTGATCGAGAGACTGGGATATTCTGCTGCGGTGCGTGGTGAAGTCAAGAT 481
DB 365 CGTCGCATGTGATCGAGAGACTGGGATATTCTGCTGCGGTGCGTGGTGAAGTCAAGAT 424

QY 482 CCACCTGGGGGGCAAGATAACTGCGAGTATCAAAATGGGGGAGCCCGCTGTCACTGAAGG 541
DB 425 CCACCTGGGGGGCAAGATAACTGCGAGTATCAAAATGGGGGAGCCCGCTGTCACTGAAGG 484

QY 542 AATTTTTCGGCGGAAATGGAAGGATGACACAAATCAATTTCTTGGCGGTGAGGTG 601
DB 485 AATTTTTCGGCGGAAATGGAAGGATGACACAAATCAATTTCTTGGCGGTGAGGTG 544

QY 602 TTGCCATTTGGCGTCCAAAGGAACTCTTGTGTACCCCTGTCAGAGTGTACGACAAAGA 661
DB 545 TTGCCATTTGGCGTCCAAAGGAACTCTTGTGTACCCCTGTCAGAGTGTACGACAAAGA 604

QY 662 AGCAAGTTTTTCAGATCTTCTACTCGGAAGACGAGGGCAAGACGTTGGAAGTTTGGGG 721
DB 605 AGCAAGTTTTTCAGATCTTCTACTCGGAAGACGAGGGCAAGACGTTGGAAGTTTGGGA 664

QY 722 AGGTAGGAGTGAATTTTGGCTGCTCTGAACCTGTGCGCTTGTAGTGGAGGGGAGCTCA 781
DB 665 AGGTAGGAGGCTTTTGGCTGCTCTGAACCTGTGCGCTTGTAGTGGAGGGGAGCTCA 724

QY 782 TCATAAACAAGTGTGACTATCGCGCGCTGTGGTGTACAGTTCAGATGACATGGGA 841
DB 725 TCATAAACAAGTGTGACTATCGCGCGCTGTGGTGTACAGTTCAGATGACATGGGA 784

QY 842 ATTCGTGGGTGAGGCTGTGCGACGCTCTCACCGTGTGTGGGGGCCCTTCACCAAAATCGA 901
```

```
DB 785 ATTCGTGGCTGGAGGCTGTGCGGACGCTCTCACTGTGTGGGGCCCCCTCACCAAAATCGA 844
QY 902 ACCAGCCCGGAGTCAGAGACGCTTCACTGCGCGTACCATCGAGGGAATCGGTGTTATGC 961
DB 845 ACCAGCCCGGAGTCAGAGACGCTTCACTGCGCGTACCATCGAGGGAATCGGTGTTATGC 904
QY 962 TCTTTCACACACCCGCTGAATTTTAAAGGGAAGGTGGTGGCGGACCCGACCTCTGGC 1021
DB 905 TCTTTCACACACCCGCTGAATTTTAAAGGGAAGGTGGTGGCGGACCCGACCTCTGGC 964
QY 1022 TGACGGATACACGAGCCATTTATAAGCTTGGGCAAGTATCCATTTGGTGTGATAAATTCG 1081
DB 965 TGACGGATACACGAGCCATTTATAAGCTTGGGCAAGTATCCATTTGGTGTGATAAATTCG 1024
QY 1082 CCTACAGCTCCGCTCTGTGTACAAGGATGATAAGCTGTACTGTGTTGATGAGTCAACAGTA 1141
DB 1025 CCTACAGCTCCGCTCTGTGTACAAGGATGATAAGCTGTACTGTGTTGATGAGTCAACAGTA 1084
QY 1142 ACGAGTGTACAGCTTGTGTTTGGCGGCTGCTGTTGGCGAGCTACGGATCATTAATCAG 1201
DB 1085 ACGAGTGTACAGCTTGTGTTTGGCGGCTGCTGTTGGCGAGCTACGGATCATTAATCAG 1144
QY 1202 TGTCTCAGTCTCTGGAAGATTCGGACACGCCACTGTCCAGCATTTGCAACCTCTGCTGATC 1261
DB 1145 TGTCTCAGTCTCTGGAAGATTCGGACACGCCACTGTCCAGCATTTGCAACCTCTGCTGATC 1204
QY 1262 CAGCGCTTCTGCTGTGAGAGCGTGTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1321
DB 1205 CAGCGCTTCTGCTGTGAGAGCGTGTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1264
QY 1322 GCTTTTGTGCGACAGTCCGCCAACCCAAACCGAATGGGAGGATGCGTACCGCTGCGTCAACG 1381
DB 1265 GCTTTTGTGCGACAGTCCGCCAACCCAAACCGAATGGGAGGATGCGTACCGCTGCGTCAACG 1324
QY 1382 CAAAGCAACGCAAAATGCGGAGAGGTTCCGAAACGGTTTGAAGTTTGGGGGTTGCGCGAG 1441
DB 1325 CAAAGCAACGCAAAATGCGGAGAGGTTCCGAAACGGTTTGAAGTTTGGGGGTTGCGCGAG 1384
QY 1442 GGGCGCTTGGCGGTCAGCCAGCAGGCGGCGAATCAACGGTATCACTTTGCAAAACAGC 1501
DB 1385 GGGCGCTTGGCGGTCAGCCAGCAGGCGGCGAATCAACGGTATCACTTTGCAAAACAGC 1444
QY 1502 CGTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1561
DB 1445 CGTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1504
QY 1562 TGGGTGCGAGCTGGACTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1621
DB 1505 TGGGTGCGAGCTGGACTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1564
QY 1622 AGCACAGTGGCAGCCAAATATACGGATCAACGCGGTCAGCGCGGATCGTGGGAGA 1681
DB 1565 GGCACAGTGGCAGCCAAATATACGGATCAACGCGGTCAGCGCGGATCGTGGGAGA 1624
QY 1682 TGGGTGAAGGTACCACTGCTTACGATGCGAATAAAATTTGCTGCTGCTGCTGCTGCTGCTG 1741
DB 1625 TGGGTGAAGGTACCACTGCTTACGATGCGAATAAAATTTGCTGCTGCTGCTGCTGCTGCTG 1684
QY 1742 ATGAGAACCTCTGAGGGTTCAGGCGACACCGTGTGTCAGACGCGGAGGAGCGCTGACA 1801
DB 1685 ATGAGAACCTCTGAGGGTTCAGGCGACACCGTGTGTCAGACGCGGAGGAGCGCTGACA 1744
QY 1802 TCTCCCACTTCTAAGTGGCGGTATGGAAGGATGATATGCAACCATTAAGCCACAGTGA 1861
DB 1745 TCTCCCACTTCTAAGTGGCGGTATGGAAGGATGATATGCAACCATTAAGCCACAGTGA 1804
QY 1862 CGGTGAATTAATCTTCTTCTTCAACCGTCACTGAGTGCAGGAGATCAGGACCTTGT 1921
DB 1805 CGGTGAATTAATCTTCTTCTTCAACCGTCACTGAGTGCAGGAGATCAGGACCTTGT 1864
QY 1922 TCTTGGCGAGGACCTGATTTGGCAGGAGCACATGCGGAGCAGCAGCGGAGCAGCTG 1981
```

Db 1865 TCTTGACCCAGGACCTGATTGGCAACGAAGACACATGTGACAGCAGCGACACGAGTG 1924

Qy 1982 AAAGA 1986

Db 1925 CCTGA 1929

RESULT 5

AR261980

LOCUS AR261980 1929 bp mRNA linear PAT 29-JAN-2003

DEFINITION Sequence 3 from patent US 6323008.

ACCESSION AR261980

VERSION AR261980.1 GI:28073278

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1929)

AUTHORS Pelletier, M., Barker, W.A., Hakes, D.J. and Zopf, D.A.

TITLE Methods for producing sialyloligosaccharides in a dairy source

JOURNAL Patent: US 6323008-A 3 27-NOV-2001;

FEATURES

Location/Qualifiers

1..1929

source /organism="unknown"

/mol_type="mRNA"

ORIGIN

Query Match 93.9%; Score 1886.6; DB 6; Length 1929;

Best Local Similarity 98.8%; Pred. No. 0;

Matches 1901; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Qy 62 TGGCACCCGGATCGAGCCGAGTTGAGCTGTTTAAAGCGGCAAAAGCTCGAAGTGCCATTG 121

Db 5 TGGCACCCGGATCGAGCCGAGTTGAGCTGTTTAAAGCGGCAAAAGCTCGAAGTGCCATTG 64

Qy 122 AAAAGGCGGCAAGTCAACGAGCGGTTGTCCACTGCTCCGCTCCCGCCCTTGTTA 181

Db 65 AAAAGGACGCAAGTCAACGAGCGGTTGTCCACTGCTCCGCTCCCGCCCTTGTTA 124

Qy 182 ATGTGGACGGGTGATGCTGTTGCCATCGGACGCTCGCTACGAAACATCAATGACAAC 241

Db 125 ATGTGGACGGGTGATGCTGTTGCCATCGGACGCTCGCTACGAAACATCAATGACAAC 184

Qy 242 CCCTCATTTGATACGGTGGCAAGTACAGCGTGGAGCGATGGGAGACCGTGGAGACCCAAA 301

Db 185 CCCTCATTTGATACGGTGGCAAGTACAGCGTGGAGCGATGGGAGACCGTGGAGACCCAAA 244

Qy 302 TTGCCATCAAGAACAGTCGTGTCATCGTCTGTTCTCGTGTGGTGGATCCCAAGTGATTG 361

Db 245 TTGCCATCAAGAACAGTCGTGTCATCGTCTGTTCTCGTGTGGTGGATCCCAAGTGATTG 304

Qy 362 TGAAGGCAACAGCTTTAGCTCCTGTTGGAGCTACACAGTTCCGAGGAGCTACTGGA 421

Db 305 TGAAGGCAACAGCTTTAGCTCCTGTTGGAGCTACACAGTTCCGAGGAGCTACTGGA 364

Qy 422 CGTCGATGTTGATCGAGAGACCTGGGATATCTGCTTGCCTGCTGGTGGTGGTCAACGAGT 481

Db 365 CGTCGATGTTGATCGAGAGACCTGGGATATCTGCTTGCCTGCTGGTGGTGGTCAACGAGT 424

Qy 482 CCACTGCGGCGGCAAGATAAATCTCGAGTATCAAAATGGGGGAGCCCGTGTCTACTGAAG 541

Db 425 CCACTGCGGCGGCAAGATAAATCTCGAGTATCAAAATGGGGGAGCCCGTGTCTACTGAAG 484

Qy 542 AATTTTCCCGGCGGCAAGATAAATCTCGAGTATCAAAATGGGGGAGCCCGTGTCTACTGAAG 601

Db 485 AATTTTCCCGGCGGCAAGATAAATCTCGAGTATCAAAATGGGGGAGCCCGTGTCTACTGAAG 544

Qy 602 TTGCCATTGTTGGGCTCAACCGGAATCTTGTGTACCTGTGCGAGTTTACGAAACAAAAGA 661

Db 545 TTGCCATTGTTGGGCTCAACCGGAATCTTGTGTACCTGTGCGAGTTTACGAAACAAAAGA 604

Qy 662 AGCAAGTTTTTCCAGATCTTCTACTCGGAAGACGAGGCGCAAGACGTTGGGG 721

Db 605 AGCAAGTTTTTCCAGATCTTCTACTCGGAAGACGAGGGCAAGACGTTGGAGTTTGGGA 664

Qy 722 AGGTAGGAGTGAATTTTGGCTGCTCTGAACTCTGTGGCCCTTTGAGTGGGAGGGAGCTCA 781

Db 665 AGGTAGGAGCGCTTTTGGCTGCTCTGAACTCTGTGGCCCTTTGAGTGGGAGGGAGCTCA 724

Qy 782 TCATAAACACTCGAGTTGACTATCGCCCGCTCTGTGTGAGTGTACGAGTCAAGTATGGGA 841

Db 725 TCATAAACACTCGAGTTGACTATCGCCCGCTCTGTGTGAGTGTACGAGTCAAGTATGGGA 784

Qy 842 ATTCTGGTGGAGGCTGTGCGACGCTCTCACTGTGTGGGCGCCCTCAACCAAAATCGA 901

Db 785 ATTCTGGTGGAGGCTGTGCGACGCTCTCACTGTGTGGGCGCCCTCAACCAAAATCGA 844

Qy 902 ACCAGCCCGGAGTCAAGAGCAGCTTCACTGCGGTGACCATCGAGGGGAATCGTGTATTGC 961

Db 845 ACCAGCCCGGAGTCAAGAGCAGCTTCACTGCGGTGACCATCGAGGGGAATCGTGTATTGC 904

Qy 962 TCTTCAACACCCCGCTGAATTTTAAAGGAAGTGGCTGCGCGACCGGACTGAACCTCTGGC 1021

Db 905 TCTTCAACACCCCGCTGAATTTTAAAGGAAGTGGCTGCGCGACCGGACTGAACCTCTGGC 964

Qy 1022 TGACGGATAACACGCGCATTTTAAAGTGGGCAAGTATCCATTTGGTGTGATGAAATTCG 1081

Db 965 TGACGGATAACACGCGCATTTTAAAGTGGGCAAGTATCCATTTGGTGTGAAATTCG 1024

Qy 1082 CCTACAGCTCCGCTCTGTACAAGGATGATTAAGCTGTACTGTTTGCATGAGATCAACAGTA 1141

Db 1025 CCTACAGCTCCGCTCTGTACAAGGATGATTAAGCTGTACTGTTTGCATGAGATCAACAGTA 1084

Qy 1142 ACAGGTGTACAGCTTGTGTTTGGCGCTGTTGGCGAGCTACGGATCATTAATCAG 1201

Db 1085 ACAGGTGTACAGCTTGTGTTTGGCGCTGTTGGCGAGCTACGGATCATTAATCAG 1144

Qy 1202 TGTGTCAGTCTCTGAAAGATTTGGGACAGCCACTGTCACGATTTGACCCCTGCTGATC 1261

Db 1145 TGTGTCAGTCTCTGAAAGATTTGGGACAGCCACTGTCACGATTTGACCCCTGCTGATC 1204

Qy 1262 CAGCCGCTTCTGTCAGAGCTGTTGTTGTCCTGTCACGAGCTTGGCTGTTGTTG 1321

Db 1205 CAGCCGCTTCTGTCAGAGCTGTTGTTGTCCTGTCACGAGCTTGGCTGTTGTTG 1264

Qy 1322 GCTTTTGTGTCAGTGTCCACCAAAACCGAATGGGAGGATGGGTACCGTGTGTTCAAG 1381

Db 1265 GCTTTTGTGTCAGTGTCCACCAAAACCGAATGGGAGGATGGGTACCGTGTGTTCAAG 1324

Qy 1382 CAGACGCGCAATGGGAGGCTTCCGAAACGTTTGAAGTTTGGGGGTTGGCGGAG 1441

Db 1325 CAGACGCGCAATGGGAGGCTTCCGAAACGTTTGAAGTTTGGGGGTTGGCGGAG 1384

Qy 1442 GGGCGCTTTGGCCGCTGAGCCAGCGGCGAGATCAACGCTATCACTTTGCAAAACCAAG 1501

Db 1385 GGGCGCTTTGGCCGCTGAGCCAGCGGCGAGATCAACGCTATCACTTTGCAAAACCAAG 1444

Qy 1502 CGTTCAAGTGTGGCTCGGTGACGATTCACAGGTTCCGAGCTCGGAGTCTTTTGC 1561

Db 1445 CGTTCAAGTGTGGCTCGGTGACGATTCACAGGTTCCGAGCTCGGAGTCTTTTGC 1504

Qy 1562 TGGGTCGAGCTCGGACTCTTCTGTGGCAAAACCTCTGCGGCTCTCGTACGACGAGA 1621

Db 1505 TGGGTCGAGCTCGGACTCTTCTGTGGCAAAACCTCTGCGGCTCTCGTACGACGAGA 1564

Qy 1622 AGCACAGTGGGAGCAATATACGATCAACGCGGCTGAGCCGAGCCGAGTCTGGGAGA 1681

Db 1565 GGCACAGTGGGAGCCATATACGATCAACGCGGCTGAGCCGAGCCGAGTCTGGGAGA 1624

Qy 1682 TGGGTAAGAGGTACCAAGTGTCTTACGATGGGGAATTAATTTGGTTCGGTGTACATTG 1741

Db 1625 TGGGTAAGAGGTACCAAGTGTCTTACGATGGGGAATTAATTTGGTTCGGTGTACATTG 1684

Qy 1742 ATGAGAACCTCTGGAGGTTTCAAGGAGACCTTGTGTCAGAGCGGAGGAGCCCTGACA 1801

Db 1685 ATGAGAACCTCTGGAGGTTTCAAGGAGACCTTGTGTCAGAGCGAGAGAGCCCTGACA 1744

QY 1802 TCTCCCACTTCTAGTTGGCGGTATGGAAGGAGTGATATGCAACATAAGCCACGTGA 1861
 Db 1745 TCTCCCACTTCTAGTTGGCGGTATGGAAGGAGTGATATGCAACATAAGCCACGTGA 1804
 QY 1862 CGGTGAATAATGTTCTTCTTTTCAACCGTCAGCTGAATGCCAGGAGATCAGGACCTTGT 1921
 Db 1805 CGGTGAATAATGTTCTTCTTTTCAACCGTCAGCTGAATGCCAGGAGATCAGGACCTTGT 1864
 QY 1922 TCTTGAGCCAGGACCTGATTTGGCA CGGAAGCACACATGGGCGACGACGCGCAGCAGTG 1981
 Db 1865 TCTTGAGCCAGGACCTGATTTGGCA CGGAAGCACACATGGGCGACGACGCGCAGCAGTG 1924
 QY 1982 AAAGA 1986
 Db 1925 CCTGA 1929

RESULT 6
 TRBTRISB
 LOCUS
 DEFINITION Trypanosoma cruzi (clone TcTS 3.5) trans-sialidase gene, complete cds.
 ACCESSION L38457
 VERSION L38457.1 GI:642914
 KEYWORDS trans-sialidase.
 SOURCE Trypanosoma cruzi
 ORGANISM Trypanosoma cruzi
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma; Schizotrypanum.
 1. (bases 1 to 1929)
 Cremona, M.I., Sanchez, D.O., Frasch, A.C. and Campetella, O.
 A single tyrosine differentiates active and inactive Trypanosoma cruzi trans-sialidases
 JOURNAL Gene 160 (1), 123-128 (1995)
 MEDLINE 95354943
 PUBMED 7628705

FEATURES
 Location/Qualifiers
 1..1929

/organism="Trypanosoma cruzi"
 /mol_type="genomic DNA"
 /strain="Tulahuen"
 /isolate="Tul2"
 /db_xref="taxon:5693"
 /clone="TcTS 3.5"
 /dev_stage="trypomastigote"
 1..1929
 /function="inactive trans-sialidase"
 /codon_start=1
 /product="trans-sialidase"
 /protein_id="AA99443.1"
 /db_xref="GI:642915"
 /translation="MLAPGSSRVELFKRQSKVPFKDKGKTVRVVHSPRLPALVNVND
 GVMVADARVETSDNSLIDTVAKYSDVDDGETWETQIAIKNSRASSVSRVVDPTVIV
 KGNKLYLVGSYNSRSYMTSHGDARDWDILLAVGVTKSTAGKKTASIKWGSVPVL
 KEFPFAEMGHTNQFLGGAGVAIVASNGNLVYPQVTKRQVPSKFIYSDDSKTW
 KFGKGRDFGCPVLALEWEGKLIINTRVDVRRRLVYESDDMNSWVEAVGFTLSRVG
 PPSKNQPSQSFTAVTIEGRKVMFLTHPLFKGRWLDRNLMLTDNRLYNVGV
 SIGDENSRLVKDDKLYLCHEINSVEYSLVFLSHATKTEMEDAYRCVNSTANAEHV
 LSSICTPADPAPSSRRGCPAVTTVGLVFLSHATKTEMEDAYRCVNSTANAEHV
 PNLKFPAGVGGALWPVQQGQNRQYEFANHALTLVASVTIHEVPRASPLLGSLDS
 SGKGLLGLSYDEKHKWQPIYGSTPVTPTGSMETGKRVHVLTMANKIGSVYIDGEP
 EGSQQTWPDPETPDISHFVYVGYKRSMDPTISHVTNNVLLYNQLNAEIRLTFLS
 QDLIGTEAHMGSSSDTSA"

ORIGIN

Query Match 93.6%; Score 1881.8; DB 3; Length 1929;
 Best Local Similarity 98.6%; Pred. No. 0;
 Matches 1898; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
 QY 62 TGGCACCCGATCGAGCCAGTTGAGCTTTTAAAGCGGCAAGCTCGAAGGTGCCATTG 121
 Db 5 TGGCACCCGATCGAGCCAGTTGAGCTTTTAAAGCGGCAAGCTCGAAGGTGCCATTG 64

QY 122 AAAAGSGCGCAAGTCAACGAGCGGGTGTCTCACTCGTTCCGGCTCCCGCCCTTGT 181
 Db 65 AAAAGGACGCAAGTCAACGAGCGGGTGTCTCACTCGTTCCGGCTCCCGCCCTTGT 124
 QY 182 ATGTGGACGGGGTGATGGTTGCCATCGCGACGCTGGCTACGAAACATCNAATGACAAC 241
 Db 125 ATGTGGACGGGGTGATGGTTGCCATCGCGACGCTGGCTACGAAACATCNAATGACAAC 184
 QY 242 CCCTCATTTGATACGGTGGCGAAGTACAGCGTGAGCGATGGGGAGACGTGGAGACCCAAA 301
 Db 185 CCCTCATTTGATACGGTGGCGAAGTACAGCGTGAGCGATGGGGAGACGTGGAGACCCAAA 244
 QY 302 TTGCCATCAAGAACAGTCGTGCATCGTCTGTCTTCTCGTGTGGTATCCACAGTCA 361
 Db 245 TTGCCATCAAGAACAGTCGTGCATCGTCTGTCTTCTCGTGTGGTATCCACAGTCA 304
 QY 362 TGAAGGCAACAAGCTTTAGTCTCGTTGGAGCTACAAAGTTTCAGGAGCTACTGGA 421
 Db 305 TGAAGGCAACAAGCTTTAGTCTCGTTGGAGCTACAAAGTTTCAGGAGCTACTGGA 364
 QY 422 CGTCGCATGGTATGCGAGAGACTGGGATATTTCTGCTTCCGCTGGTGGAGTCAAGACT 481
 Db 365 CGTCGCATGGTATGCGAGAGACTGGGATATTTCTGCTTCCGCTGGTGGAGTCAAGACT 424
 QY 482 CCACTCGGGCGGCGAAGATACTGCGAGTATCAAAATGGGGAGCCCGCTGTCACTGAAGG 541
 Db 425 CCACTCGGGCGGCGAAGATACTGCGAGTATCAAAATGGGGAGCCCGCTGTCACTGAAGG 484
 QY 542 AATTTTCCCGCGGGAATGGAGAAATGCACACAAATCAATTTCTTGGCGGTGCAAGTG 601
 Db 485 AATTTTCCCGCGGGAATGGAGAAATGCACACAAATCAATTTCTTGGCGGTGCAAGTG 544
 QY 602 TTGCCATTTGTGGCGTCCAAACGGGAATCTTGTTGTACCTGTGCAGTTACGAAACAA 661
 Db 545 TTGCCATTTGTGGCGTCCAAACGGGAATCTTGTTGTACCTGTGCAGTTACGAAACAA 604
 QY 662 AGCAAGTTTTCGAAGATCTTTACTCGGAAGACGAGGGCAAGACGTGGAGTTGGGG 721
 Db 605 AGCAAGTTTTCGAAGATCTTTACTCGGAAGATGATGGCAAGACGTGGAGTTGGGA 664
 QY 722 AGGGTAGAGTGATTTTGGCTGCTCTGAACCTGTGGCCCTTGAAGTGGAGGGGAGCTCA 781
 Db 665 AGGGTAGAGTGATTTTGGCTGCTCTGAACCTGTGGCCCTTGAAGTGGAGGGGAGCTCA 724
 QY 782 TCATAAACACTCGAGTTGACTATCGCCCGCTGTGGTGTACGAGTCCAGTGACATGGGA 841
 Db 725 TCATAAACACTCGAGTTGACTATCGCCCGCTGTGGTGTACGAGTCCAGTGACATGGGA 784
 QY 842 ATTCTGGTGGAGGCTGTGGGACGCTCTCACGTGTGGGGCCCTCACCAAAATCGA 901
 Db 785 ATTCTGGTGGAGGCTGTGGGACGCTCTCACGTGTGGGGCCCTCACCAAAATCGA 844
 QY 902 ACCAGCCCGCAGTCAAGAGAGCTTCACTCCGTGACCATCGAGGGGAATCGTGTATGC 961
 Db 845 ACCAGCCCGCAGTCAAGAGAGCTTCACTCCGTGACCATCGAGGGGAATCGTGTATGC 904
 QY 962 TCTTCAACACCCGCTCAATTTTAAAGGAAGGTGGTGGCGGACCGACTGAACCTCTGGC 1021
 Db 905 TCTTCAACACCCGCTCAATTTTAAAGGAAGGTGGTGGCGGACCGACTGAACCTCTGGC 964
 QY 1022 TGAAGGATAACAGCGCATTTATAAGTTGGGCAAGTATCCATTTGGTGTGAAATTTCCG 1081
 Db 965 TGAAGGATAACAGCGCATTTATAAGTTGGGCAAGTATCCATTTGGTGTGAAATTTCCG 1024
 QY 1082 CCTACAGTCCGTCCTGTACAAGATGATAGCTGTACTGTGTGATGAGATCAACAGTA 1141
 Db 1025 CCTACAGTCCGTCCTGTACAAGATGATAGCTGTACTGTGTGATGAGATCAACAGTA 1084
 QY 1142 ACCAGGTGTACAGCTTGTGTGTGGCGGCTGTGTGGCGAGCTACGGAATCATTAATCAG 1201
 Db 1085 ACCAGGTGTACAGCTTGTGTGTGGCGGCTGTGTGGCGAGCTACGGAATCATTAATCAG 1144

Qy	1202	TGCTGCAGTCTCTGGAAGAAATTGGGACAGCCACCTGTCCAGCATTTGACCCCTGCTGATC	1261
Db	1145	TGCTGCAGTCTCTGGAAGAAATTGGGACAGCCACCTGTCCAGCATTTGACCCCTGCTGATC	1204
Qy	1262	CAGCCGCTTCGTCTGTCAGAGCGTGTGTGTCCTCCGCTGTCCACACGTTGCTCTGTTG	1321
Db	1205	CAGCCGCTTCGTCTGTCAGAGCGTGTGTGTCCTCCGCTGTCCACACGTTGCTCTGTTG	1264
Qy	1322	GCTTTTGTGTCGACAGTGCACCAAAACCAGAAATGGGAGGATGCGTACCGTCCGCTCAACG	1381
Db	1265	GCTTTTGTGTCGACAGTGCACCAAAACCAGAAATGGGAGGATGCGTACCGTCCGCTCAACG	1324
Qy	1382	CACACCGGCAAAATCCGAGAGGTTCCGAACGTTTGAAGTTTGGCGGGTGGCGGAG	1441
Db	1325	CAAGCACGGCAAAATCCGAGAGGTTCCGAACGTTTGAAGTTTGGCGGGTGGCGGAG	1384
Qy	1442	GGCGCTTTTCGCGGTGAGCGAGCGAGCGGGCAGAAATCAACGTTATCATTTGCCAAACCAACG	1501
Db	1385	GGCGCTTTTCGCGGTGAGCGAGCGAGCGGGCAGAAATCAACGTTATGATTTGCCAAACCAACG	1444
Qy	1502	CGTTACGCTGTGGCGTCCGTCGATGACGATTCACGAGGTTCCGAGCGTCGCGAGTCTCTTTC	1561
Db	1445	CGTTACGCTGTGGCGTCCGTCGATGACGATTCACGAGGTTCCGAGCGTCGCGAGTCTCTTTC	1504
Qy	1562	TGGTGGACCTTGGACCTCTTCTGCTGGCAAAAATCTCTGGGCTCTCTGTCGACGAGA	1621
Db	1505	TGGTGGACCTTGGACCTCTTCTGCTGGCAAAAATCTCTGGGCTCTCTGTCGACGAGA	1564
Qy	1622	AGCACAGTGGCAGCCAAATATACGATCAACGCGGTGACGCCGACCGGATCGTGGAGA	1681
Db	1565	AGCACAGTGGCAGCCAAATATACGATCAACGCGGTGACGCCGACCGGATCGTGGAGA	1624
Qy	1682	TGGTAAGAGGTACACGCTGTTCTTACGATGCGAATAAAATGTTCCGTTGTCATATG	1741
Db	1625	CGGTAAGAGGTACACGCTGTTCTTACGATGCGAATAAAATGTTCCGTTGTCATATG	1684
Qy	1742	ATGAGAACCTCTGAGGGTTACGGGACAGACCGTTGTGTCAGACGGGAGACCGCTGACA	1801
Db	1685	ATGAGAACCTCTGAGGGTTACGGGACAGACCGTTGTGTCAGACGGGAGACCGCTGACA	1744
Qy	1802	TCCTCCACTTCTACGTTGGCGGTATGGAAGGATGATATGCCAACCAATAGCCAGTGA	1861
Db	1745	TCCTCCACTTCTACGTTGGCGGTATGGAAGGATGATATGCCAACCAATAGCCAGTGA	1804
Qy	1862	CGGTGAATAATGTTCTTTTACAAACCGTCAGCTGATGTCGCGAGGATCAGGACCTTGT	1921
Db	1805	CGGTGAATAATGTTCTTTTACAAACCGTCAGCTGATGTCGCGAGGATCAGGACCTTGT	1864
Qy	1922	TCTTGAGCCAGGACCTGATTTGGCAGCGAAGCACACATGCGCAGACAGCGGACGAGTG	1981
Db	1865	TCTTGAGCCAGGACCTGATTTGGCAGCGAAGCACACATGCGCAGACAGCGGACGAGTG	1924
Qy	1982	AAAGA 1986	
Db	1925	CCTGA 1929	
RESULT 7			
TRBTRISIC			
LOCUS	Trypanosoma cruzi (clone TcTST4)	trans-sialidase gene, complete cds.	linear INV 14-MAY-1996
DEFINITION			
ACCESSION	L38463		
VERSION	L38463.1	GI:642916	
KEYWORDS	trans-sialidase.		
SOURCE	Trypanosoma cruzi		
ORGANISM	Trypanosoma cruzi		
	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;		
	Trypanosoma; Schizotrypanum.		
REFERENCE	1 (bases 1 to 1929)		
AUTHORS	Cremona,M.L., Sanchez,D.O., Frasch,A.C. and Campetella,O.		
TITLE	A single tyrosine differentiates active and inactive Trypanosoma cruzi trans-sialidases		

JOURNAL	Gene 160	(1), 123-128	(1995)
MEDLINE	95354943		
PUBMED	7628705		
FEATURES			
source	Location/Qualifiers		
	1..1929		
	/organism="Trypanosoma cruzi"		
	/mol_type="genomic DNA"		
	/strain="Tulahuen"		
	/isolate="Tul2"		
	/db_xref="taxon:5693"		
	/clone="TcTST4"		
	/dev_stage="trypanaestigote"		
CDS	1..1929		
	/codon_start=1		
	/product="trans-sialidase"		
	/protein_id="AAA99444.1"		
	/db_xref="GI:642917"		
	/translation="MLAPGSSRVLEPKROSSKVPFKDGVKTVRVVHSFRLPALVND		
	GVNVALADARYETSDNLSLIDTVKYSVDGGETWETQIAIKNSRASSVSRVDPVIV		
	KGNKLVLVGSYNSRSRYTSHVGDARDWNILLAVGEVTKTAGCKITASKWSPVSL		
	KEFPFAEMEGMHTNQFLGGAGVAIVASNGNLVYPVQVTKRKQVFSKIFYSBDDKTM		
	KFFGRSDFCSEPVLEWEGKLIINTRVDLRFKRWLRDLNLWLDNORIYNVGOV		
	PSPKSNQPSQSSFTAVTTEGMRVMLFTHPLNFKRWLRDLNLWLDNORIYNVGOV		
	SIGDENSAYSSVLYKDDKLYCLHEINSNEVYSLVFARLVGELRIIRKSVLQSNKWDH		
	LSISCTPADPAASRSRGCGPATTVGLVGLSHSATKTEWEDAYRCVNASMANAERV		
	PNGLKFAVGVGAGALWPVSOQGNQRYRPFANHAFTVVASVTHIEVPVSASPLGASLDS		
	SGCKLLGLSYDEKHWQPIYGSTPVTPTFPGSWETGRYHVVTLMANKIGSVYIDGEP		
	EGSGQTVWPDPERTPDISHFVGVGYSKSDMPTISHVTNNVLLNRLNABEIRTLFLS		
	ODLIGTEAHMDSSDSTA"		
ORIGIN			
	Query Match 93.4%; Score 1877; DB 3; Length 1929;		
	Best Local Similarity 98.4%; Pred. No. 0;		
	Matches 1895; Conservative 0; Mismatches 30; Indels 0; Gaps 0;		
Qy	62	TGGCACCCGATCGAGCCGAGTTGAGCTGTTTAAGCGGCAAGCTCGAAGTGCCATTG	121
Db	5	TGGCACCCGATCGAGCCGAGTTGAGCTGTTTAAGCGGCAAGCTCGAAGTGCCATTG	64
Qy	122	AAAAGCGCGCAAGTCAACGAGCGGTTGTCACCTCGTTCCGCTCCCCCGCTTGTGA	181
Db	65	AAAAGCGCGCAAGTCAACGAGCGGTTGTCACCTCGTTCCGCTCCCCCGCTTGTGA	124
Qy	182	ATGTGAACGGGTGATGTTGCCATCGCGACGCTCGCTACGAAACATCAATGACAACT	241
Db	125	ATGTGAACGGGTGATGTTGCCATCGCGACGCTCGCTACGAAACATCAATGACAACT	184
Qy	242	CCCTCATGATACGTTGGCGAAGTACAGCGTGACAGTGGGAGAGCTGGGAGACCCAAA	301
Db	185	CCCTCATGATACGTTGGCGAAGTACAGCGTGACAGTGGGAGAGCTGGGAGACCCAAA	244
Qy	302	TTGCCATCAAGAACAGTCGTGTCATCGTCTGTTCTCGTGTGGTGCATCCACAGTGAATG	361
Db	245	TTGCCATCAAGAACAGTCGTGTCATCGTCTGTTCTCGTGTGGTGCATCCACAGTGAATG	304
Qy	362	TGAAGGCGCAACAGCTTTACGTCCTCGTTGGAAAGTACAAACAGTTCGAGAGCTACTGGA	421
Db	305	TGAAGGCGCAACAGCTTTACGTCCTCGTTGGAAAGTACAAACAGTTCGAGAGCTACTGGA	364
Qy	422	CGTTCGATGTTGATCGAGAGCTGSGATATTCCTGCTTGCCTTGGTGGAGTCAACGAGT	481
Db	365	CGTTCGATGTTGATCGAGAGCTGSGATATTCCTGCTTGCCTTGGTGGAGTCAACGAGT	424
Qy	482	CCACTCGCGCGGCAAGATAACTCGAGGTATCAAAATGGGGAGCCCGCTGTCACTGAAGG	541
Db	425	CCACTCGCGCGGCAAGATAACTCGAGGTATCAAAATGGGGAGCCCGCTGTCACTGAAGG	484
Qy	542	AATTTTTCGCGCGGAAATGGAAGGAATGCACAAATCAATTTCTTGGCGGTGCAGGTG	601
Db	485	AATTTTTCGCGCGGAAATGGAAGGAATGCACAAATCAATTTCTTGGCGGTGCAGGTG	544
Qy	602	TTGCCATGTTGGCGTCCAAACGGGAATCTTGTGTACCTGTGTCAGGTTTACGAAACAAAGA	661

Db 545 TTGCCATTGTGGCGTCCAAACGGGAATCTTGTGTACCTGTGCAGGTTACGAACAAGGA 604
Qy 662 AGCAAGTTTTTTTCCAGATCTTCTACTCGGAACGAGGCGCAGAGCTGGAAGTTTGGG 721
Db 605 AGCAAGTTTTTTTCTAAGATCTTCTACTCGGAATGATGCGAAGAGCTGGAAGTTTGGGA 664
Qy 722 AGGTAGAGTGAATTTTGGCTGCTGAAACCTGTGGCCCTTGAAGTGGAGGGAAGCTCA 781
Db 665 AGGTAGAGTGAATTTTGGCTGCTGAAACCTGTGGCCCTTGAAGTGGAGGGAAGCTCA 724
Qy 782 TCATAAACACTCCAGTGTGACTATCGCGCGCGTCTGTGTACAGTCCAGTGACATGGGGA 841
Db 725 TCATAAACACTCCAGTGTGACTATCGCGCGCGTCTGTGTACAGTCCAGTGACATGGGGA 784
Qy 842 ATTCTGGGTGGAGGCTGTCCGCAACCTCTCACAGTGTGTGGGCGCCCTCACCAAAATCGA 901
Db 785 ATTCTGGGTGGAGGCTGTCCGCAACCTCTCACAGTGTGTGGGCGCCCTCACCAAAATCGA 844
Qy 902 ACCAGCCCGCAGTCAAGCAGCTTCACTGCGTGAACCATCGAGGGAATGCGTGTATGC 961
Db 845 ACCAGCCCGCAGTCAAGCAGCTTCACTGCGTGAACCATCGAGGGAATGCGTGTATGC 904
Qy 962 TCTTTCACACACCGCTGAATTTTAAGGGAAGTGGCTGGCGGACCGACTGAACCTCTGGC 1021
Db 905 TCTTTCACACACCGCTGAATTTTAAGGGAAGTGGCTGGCGGACCGACTGAACCTCTGGC 964
Qy 1022 TGAAGGATACACAGCGCAATTTATACGTTGGGCAAGTATCCATTGGTGATGAAATTCG 1081
Db 965 TGACGGATAACACAGCGCAATTTATACGTTGGGCAAGTATCCATTGGTGATGAAATTCG 1024
Qy 1082 CCTACAGTCCGCTGACAGGATGATAGCTGTACTGTGTTGCATGAGATCAACAGTA 1141
Db 1025 CCTACAGTCCGCTGACAGGATGATAGCTGTACTGTGTTGCATGAGATCAACAGTA 1084
Qy 1142 ACGAGGTGTACAGCTTGTGTTGCGCGCTGTTGGCGAGTACGATCAAGATCAATTAATCAG 1201
Db 1085 ACGAGGTGTACAGCTTGTGTTGCGCGCTGTTGGCGAGTACGATCAAGATCAATTAATCAG 1144
Qy 1202 TGCTGAGTCTGGAAGATTTGGGACAGCCACTGTCCAGCATTTGACCCCTGCTGATC 1261
Db 1145 TGCTGAGTCTGGAAGATTTGGGACAGCCACTGTCCAGCATTTGACCCCTGCTGATC 1204
Qy 1262 CAGCGCTTCTGCTCAGAGCGTGTGTGTCGCGCTGTCAGCATTTGACCCCTGCTGATC 1321
Db 1205 CAGCGCTTCTGCTCAGAGCGTGTGTGTCGCGCTGTCAGCATTTGACCCCTGCTGATC 1264
Qy 1322 GCTTTTGTGCGCACAGTGCCACCAAAACCGAATGGGAGATGCGTACCGCTGCGTGAACG 1381
Db 1265 GCTTTTGTGCGCACAGTGCCACCAAAACCGAATGGGAGATGCGTACCGCTGCGTGAACG 1324
Qy 1382 CAAGCAGCGCAAAATGCGGAGAGGTTCCGAACCGTTTGAAGTTTGGCGGGTTGGCGGAG 1441
Db 1325 CAAGCAGCGCAAAATGCGGAGAGGTTCCGAACCGTTTGAAGTTTGGCGGGTTGGCGGAG 1384
Qy 1442 GGGCGCTTTGGCGGTTGAGCAGCAGCGGGCAGAAATCAACGTTATCACTTTGCAAAACCG 1501
Db 1385 GGGCGCTTTGGCGGTTGAGCAGCAGCGGGCAGAAATCAACGTTATCACTTTGCAAAACCG 1444
Qy 1502 GCTTACCGCTGGGTGGGTGCGTGAAGATTCACGAGTTCCGAGCGTCCGAGTCCCTTTCG 1561
Db 1445 GCTTACCGCTGGGTGGGTGCGTGAAGATTCACGAGTTCCGAGCGTCCGAGTCCCTTTCG 1504
Qy 1562 TGGGTGCGAGCTTGGACTCTTCTGTTGGCAAAATCTCTGGGGCTCTGTCAGCAGAGA 1621
Db 1505 TGGGTGCGAGCTTGGACTCTTCTGTTGGCAAAATCTCTGGGGCTCTGTCAGCAGAGA 1564
Qy 1622 AGCAGAGTGGAGCAATATACGATCAACCGCGTGAACCGGACCGGATCGTGGGAGA 1681
Db 1565 AGCAGAGTGGAGCAATATACGATCAACCGCGTGAACCGGACCGGATCGTGGGAGA 1624
Qy 1682 TGGGTGAAGAGTACACAGCTGTTCTTACAGTGGCGCAATAAATTTGGTTCGGTGTACATTG 1741
Db 1625 CGGTTAAGAGTACACAGCTGTTCTTACAGTGGCGCAATAAATTTGGTTCGGTGTACATTG 1684

Qy 1742 ATGAGAACTCTGGAGGGTTTCAGGGCAGACCGTGTGTGCCAGACGGAGACCGCTGACA 1801
Db 1685 ATGAGAACTCTGGAGGGTTTCAGGGCAGACCGTGTGTGCCAGACGGAGACCGCTGACA 1744
Qy 1802 TCTCCCACTTCTACGTTGGCGGTATGGAAGAGTATATGCCAACCATTAAGCCAGTGA 1861
Db 1745 TCTCCCACTTCTACGTTGGCGGTATGGAAGAGTATATGCCAACCATTAAGCCAGTGA 1804
Qy 1862 CGTGTGAATATGTTCTTCTTACCAACCGTCAGCTGAATGCCAGGAGATCAGACCTTGT 1921
Db 1805 CGTGTGAATATGTTCTTCTTACCAACCGTCAGCTGAATGCCAGGAGATCAGACCTTGT 1864
Qy 1922 TCTTGACCGAGGACCTGATTTGGCAGCAAGACACATGTCGAGCAGCAGCAGTGA 1981
Db 1865 TCTTGACCGAGGACCTGATTTGGCAGCAAGACACATGTCGAGCAGCAGCAGTGA 1924
Qy 1982 AAAGA 1986
Db 1925 CCTGA 1929

TRBTRSTIA
Trypanosoma cruzi 1929 bp DNA linear INV 14-MAY-1996
LOCUS
DEFINITION
L38456
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
FEATURES
1. .1929
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/strain="RA"
/db_xref="taxon:5693"
/clone="TCT52V0"
/db_stage="trypanastigote"
1. .1929
/function="inactive"
/codon_start=1
/product="trans-sialidase"
/protein_id="AAA99442.1"
/db_xref="GI:642913"
/translation="MLAGSSRRVELPKRSSKVPKSGKVTERRVHSPRLPALVNVDD
GVNVAADARYETSDNLSLIDTVAKSYVDGDETWTQIAIKNSRASSVSRRVDPVIV
KGNKLVLVGSSNYSSTWTSHTGDARDWDILLAVESTKTAGKITSIKWGSPPVSL
KEFPFPAEMGHTNPLFGAGVAIVASNGNLVTPQVNTKKQVFSKIPYSEDGKTW
KFGKSAFAGCEAVALWEGKLIINTRDYRRRLVYESSDMGTWLEAVGTLRSVWG
PSPKSPQGSQSFATVITTEGRVLMITPLNFKGRLDRNLNLWNTQRIYNYGOV
SIGDENSASHSVLYKDDKLYCLHEINSNEVSYLVFARLVGLRILIKSVLOSQKNDSH
LSISICTPADPAASSSRGCGPATTVGLVPLSHSATKTEWEDAYCYNASTANAEVR
PNGLKPAVGCGALMPVSPQONQRYRFAHAFVTVASVTHEVESVAPLLGASLRS
SGGKLLGLSYDEKHQWPIYGSTPTVPTGWSWETGKTHVLTMTANKTGSVTDGEPL
CQSGQTVDPDERTPDISHFYGGYKRSMDPTTISHVTNNVNNLLYNQLNABBITRLFLS
QLLIGTEAHMDSSTISA"

source

CDS

ORIGIN

Query Match 93.2%; Score 1873.8; DB 3; Length 1929;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1893; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 62 TGGCACCCGGATCGAGCCGAGTTGAGCTGTTTAAAGCGGCAAGCTCGAAGGTGCTCATTTG 121
DB 5 TGGCACCCGGATCGAGCCGAGTTGAGCTGTTTAAAGCGGCAAGCTCGAAGGTGCTCATTTG 64
QY 122 AAAAGGCGCAAGAGTCACAGCGGCTGTCACATCGTTCCGCTCCCGCCCTTGTTA 181
DB 65 AAAAGGCGCAAGAGTCACAGCGGCTGTCACATCGTTCCGCTCCCGCCCTTGTTA 124
QY 182 ATGTGACGGGGTGATGTTGCCATCGCGAGCGCTCGCTACGAAACATCAATGACAAC 241
DB 125 ATGTGACGGGGTGATGTTGCCATCGCGAGCGCTCGCTACGAAACATCAATGACAAC 184
QY 242 CCCTCATTTGATACGTTGGCCAGTACAGCTGAGACGATGGGAGACGTTGGGAGACCCAAA 301
DB 185 CCCTCATTTGATACGTTGGCCAGTACAGCTGAGACGATGGGAGACGTTGGGAGACCCAAA 244
QY 302 TTGCCATCAAGAACAGTCGTCATCGTCTGTTCTCGTGTGGTGGATCCACAGTGAATTG 361
DB 245 TTGCCATCAAGAACAGTCGTCATCGTCTGTTCTCGTGTGGTGGATCCACAGTGAATTG 304
QY 362 TGAAGGCGCAAGAGCTTTACGCTCTGTTGGAAAGCTACAAACAGTTCGAGGAGCTACTGGA 421
DB 305 TGAAGGCGCAAGAGCTTTACGCTCTGTTGGAAAGCTACAAACAGTTCGAGGAGCTACTGGA 364
QY 422 CGTGCATGTTGATGCGAGAGATCGGATATCTGTTGCGGTTGGTGGAGTCAAGAGT 481
DB 365 CGTGCATGTTGATGCGAGAGATCGGATATCTGTTGCGGTTGGTGGAGTCAAGAGT 424
QY 482 CCACTCGGCGCGCAGATAACTCGAGTATCAAAATGGGGGAGCCCGTCTCACTGAAGG 541
DB 425 CCACTCGGCGCGCAGATAACTCGAGTATCAAAATGGGGGAGCCCGTCTCACTGAAGG 484
QY 542 AATTTTTTCCCGCGGAAATGGAAGGATGACACAAATCAATTTCTTTGGCGGTGCGAGGTG 601
DB 485 AATTTTTTCCCGCGGAAATGGAAGGATGACACAAATCAATTTCTTTGGCGGTGCGAGGTG 544
QY 602 TTGCCATTTGGCGTCAACGGGAATCTGTGTACCTGTGCGAGTTACGAAACAAAGA 661
DB 545 TTGCCATTTGGCGTCAACGGGAATCTGTGTACCTGTGCGAGTTACGAAACAAAGA 604
QY 662 AGCAAGTTTTTCCAGATCTTCTACTCGGAACGAGGCGCAAGAGTGGAACTTTGGG 721
DB 605 AGCAAGTTTTTCCAGATCTTCTACTCGGAAGATGATGGCAAGAGTGGAACTTTGGG 664
QY 722 AGGTAGGAGTGAATTTGGCTGCTCTGAACCTGTGGCCCTTGAGTGGAGGGGAAGCTCA 781
DB 665 AGGTAGGAGTGAATTTGGCTGCTCTGAAGCTGTGGCCCTTGAGTGGAGGGGAAGCTCA 724
QY 782 TCATAAACACTCGAGTTGACTATCGCGCGCGTCTGTGTACGAGTCCAGTGAATGGGGA 841
DB 725 TCATAAACACTCGAGTTGACTATCGCGCGCGTCTGTGTACGAGTCCAGTGAATGGGGA 784
QY 842 ATTCTGGTGGTGGAGGCTGTGGCAGCTCTCAGCTGTGGGCGCCCTCACCAAAATCGA 901
DB 785 ATACGTGGTGGAGGCTGTGGCAGCTCTCAGCTGTGGGCGCCCTCACCAAAATCGA 844
QY 902 ACCAGCCCGGAGTCAAGAGCTTCACTGCCGTGACCAATCGAGGGAATCGGTGTTATGC 961
DB 845 ACCAGCCCGGAGTCAAGAGCTTCACTGCCGTGACCAATCGAGGGAATCGGTGTTATGC 904
QY 962 TCTTTCACACCCCGTGAATTTTAAAGGAAGTGGCTGGCGGACCGAGTGAACCTCTGGC 1021
DB 905 TCTTTCACACCCCGTGAATTTTAAAGGAAGTGGCTGGCGGACCGAGTGAACCTCTGGC 964
QY 1022 TGAAGGATACCAAGCCATTTATAAGTTGGGCAAGTATCCATGTTGATGAATAATCCG 1081
DB 965 TGAAGGATACCAAGCCATTTATAAGTTGGGCAAGTATCCATGTTGATGAATAATCCG 1024
QY 1082 CCTTACAGCTCCGTCCTGTACAAGGATGATGAAGCTGTTGTTGATGAGTCAACAGTA 1141
DB 1025 CCCACAGCTCCGTCCTGTACAGAGATGATGAAGCTGTTGTTGATGAGTCAACAGTA 1084
QY 1142 ACGAGGTGTACAGCTTGTGTTTGGCGGCGCTGTTGGCGAGCTACGGATCATTTAAATCAG 1201

DB 1085 ACAGGTGTACAGCTTGTGTTTGGCGGCTGTTGGCGAGCTACGGATCATTTAAATCAG 1144
QY 1202 TGTGTCAGTCTCTGGAAGAAATTTGGGACAGCCACCTGTCCAGCATTTGACCCCTGCTCATC 1261
DB 1145 TGTGTCAGTCTCTGGAAGAAATTTGGGACAGCCACCTGTCCAGCATTTGACCCCTGCTCATC 1204
QY 1262 CAGCCGCTTGTGTCGTGTCAGAGCGTGGTGTGCTCCGCTGTCAACACGGTGTGTTGTTG 1321
DB 1205 CAGCCGCTTGTGTCGTGTCAGAGCGTGGTGTGCTCCGCTGTCAACACGGTGTGTTGTTG 1264
QY 1322 GCTTTTGTGTCAGACAGTGCACCAAAACCGAAATGGGAGGATGCTACCGCTGGTGAACG 1381
DB 1265 GCTTTTGTGTCAGACAGTGCACCAAAACCGAAATGGGAGGATGCTACCGCTGGTGAACG 1324
QY 1382 CAAGCACGGCAAAATCGGAGAGGGTTCGGAACCGTTCGAAAGTTTGGCGGGTGGCGGAG 1441
DB 1325 CAAGCACGGCAAAATCGGAGAGGGTTCGGAACCGTTCGAAAGTTTGGCGGGTGGCGGAG 1384
QY 1442 GGGCGCTTTGGCGGCTGAGCCAGCAGGGGAGAAATCAACGGTATCACTTTGCAAAACACG 1501
DB 1385 GGGCGCTTTGGCGGCTGAGCCAGCAGGGGAGAAATCAACGGTATCGCTTTGCAAAACACG 1444
QY 1502 CGTTCAGCTTGGTGGCGTGGTGAAGTTTCAAGAGTTTCCGAGGTCGCGAGTCTTTTGC 1561
DB 1445 CGTTCAGCTTGGTGGCGTGGTGAAGTTTCAAGAGTTTCCGAGTTCGCGAGTCTTTTGC 1504
QY 1562 TGGGTGCGAGCTTGGACTCTTCTGTGTGGCAAAAACTCTGTGGGCTCTCTGTAACGAGGA 1621
DB 1505 TGGGTGCGAGCTTGGACTCTTCTGTGTGGCAAAAACTCTGTGGGCTCTCTGTAACGAGGA 1564
QY 1622 AGCACAGTGGCAGCCAAATATACGATCAACCGCGGTGACGCCGACCGGATCGTGGGAGA 1681
DB 1565 AGCACAGTGGCAGCCAAATATACGATCAACCGCGGTGACGCCGACCGGATCGTGGGAGA 1624
QY 1682 TGGGTAAAGGTACCAAGTGGTTCCTACGATGGCGAATAAAATGGTTCGGTGTACATTTG 1741
DB 1625 CGGTTAAAGGTACCAAGTGGTTCCTACGATGGCGAATAAAATGGTTCGGTGTACATTTG 1684
QY 1742 ATGCGAAGCTCTGAGAGGTTTCAAGGCGAGACCGTGTGTCAGACGGGAGGACCGCTGACA 1801
DB 1685 ATGCGAAGCTCTGAGAGGTTTCAAGGCGAGACCGTGTGTCAGACGGGAGGACCGCTGACA 1744
QY 1802 TCTCCACTTCTACGTTGGCGGATGGAAGGATGATATGCAACATTAAGCAACGTGA 1861
DB 1745 TCTCCACTTCTACGTTGGCGGATGGAAGGATGATATGCAACATTAAGCAACGTGA 1804
QY 1862 CGGTGAATATGTTCTTCTTACACCGTCAAGTGAATGCCGAGGAGATCAGGACCTTGT 1921
DB 1805 CGGTGAATATGTTCTTCTTACACCGTCAAGTGAATGCCGAGGAGATCAGGACCTTGT 1864
QY 1922 TCTTGAAGCAGGACCTGATTTGGCAGGAAAGCACATATGGCAGCAGCAGCGGAGCATGTG 1981
DB 1865 TCTTGAAGCAGGACCTGATTTGGCAGGAAAGCACATATGGCAGCAGCAGCGGAGCATGTG 1924
QY 1982 AAAGA 1986
DB 1925 CTTGA 1929

RESULT 9
AF525766/c

LOCUS Trypanosoma cruzi 30095 bp DNA linear INV 02-JUL-2003
DEFINITION Trypanosoma cruzi unknown genes.
ACCESSION AF525766
VERSION AF525766.1 GI:32401133
KEYWORDS
SOURCE Trypanosoma cruzi
ORGANISM Trypanosoma cruzi
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
REFERENCE 1 (bases 1 to 30095)
AUTHORS Kim,D.H., Santos,M.R.M., Araya,J.E., Souza,R.T., Gomes,H.B.M.,

TITLE
JOURNAL
Direct Submission
Submitted (20-Jun-2002) Microbiologia Immunologia e Parasitologia,
Universidade Federal de Sao Paulo - UNIFESP/EPM, Botucatu 862, Sao
Paulo, SP 04023-062, Brasil
FEATURES
source
1..30095
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/strain="Sylvio"
/db_xref="taxon:5693"
/clone="X10/6"
/complement(<1158..>2117)
/product="unknown"
/complement(1158..2117)
/note="TcC2-ORF1; similar to Drosophila melanogaster
CG12017 gene product encoded by GenBank Accession Number
AB003477; also similar to Trypanosoma cruzi EST clones
TENS2302, TENS2422, TENS2425, and TENS1846 deposited in
GenBank Accession Numbers AI562310, AI562776, AI562778,
and AA952662"
/codon_start=1
/product="unknown"
/protein_id="AAP80759.1"
/db_xref="GI:32401134"
/translation="MAMMTGRVLLVLCALCVLWCGARGGFADEREGLSGGBPLIQS
KALEKPLQVQIRDEAVPLKELPAPAGDEDEDDDGNGAGCTHEEKPIGRQ
GKGQTTVPQDSTETNLGKQQTGOSIVSAGDISHSGSOEAILTQPTKQNET
HKNTPAVENLATVNGENTLLAGIAGNLPSPEDGVDSEHQSGEDITSEDEKNVSP
ETAATPQSHREGSEGTEGKGKATIVAAATTTNTTNSDSSTVKMSEAAQTITLT
AAQTNHTVPGDSGSAVSHSTSPLLLILLVACAAAAVAA"
/product="unknown"
/complement(4031..>4885)
/product="unknown"
/note="TcC2-ORF2; similar to Trypanosoma cruzi surface
glycoprotein GP90 encoded by GenBank Accession Number
AF426132; also similar to Trypanosoma cruzi EST clones
TENS1965, TEUP0261, TENG0653, and TENF0415 deposited in
GenBank Accession Numbers AI562437, AA426681, AI664666,
and AA556107"
/codon_start=1
/product="unknown"
/protein_id="AAP80760.1"
/db_xref="GI:32401135"
/translation="MVRISCSAPCCARMASRLLRQSGSGRVSLSRLMELSAL
RSLVSTWAQDIIPFSLSIHTAGLVLSDAASDDTWDEYRCLYATVNAVKDNYGL
QVTGIESRAIWPVNTSRDEVRVLSHNTFTLVASVSIETPQSGNTPLTAMLADTASN
HNMGLSYSHNKWETPEGKTTTSCSANEPRKEYQVALMLQGNKASVDVDSGLGEE
VPLTGERPLEVRLCFGACGCGSHSVTVDLFLHNRPLNSTARRALKDRAPRGRRL
RHHSKRE"
/complement(<7929..>10487)
/product="unknown"
/complement(7929..10487)
/note="TcC2-ORF3; similar to Trypanosoma cruzi surface
proteins deposited in PIR Accession Number T30521 and
encoded by GenBank Accession Numbers AB010287 and
AB017765; also similar to Trypanosoma cruzi EST clones
TENU0307, TENG0947, TENS0203, and TENS0222 deposited in
GenBank Accession Numbers AA883049, AI668025, AA882632,
and AI144563"
/codon_start=1
/product="unknown"
/protein_id="AAP80761.1"
/db_xref="GI:32401136"
/translation="MNGDVFVAEAOCKKNECVFTGIASEFLSLTDEKPKGLDKSL
KTQVLGNRSGGKCPQSOTAVLPQTRVKLYVRPPTVNGSHIYMLAGNYWTYQD
NADVPQWGLLVKGNISGEENKRIWKDYDLPTWSNGRQESWTQLIGGGSGGI
KLKGGLLPGLRTCKGKAVSLTWSSDTGEGTVSRMSDGGSDPSVWEGEKD
KIMMGTTACDGRVRVYESGDGMSTEALGTLRSRWNNKKGAKGVRSGFIITADGVE
DNRTNVALPYAKTESENVEKEKEKRLHLMTDNTHTLVDIGPVTSKDDDAVAS
SLLYSKTNGKELTALYEKKDGDDEASPGMVSVLLTEQLKRVKEVLTWKEVDDRV
FKUCPSSESPLSENACSTFTFKVTDGLVGLSDSFSDTDTRDEYLVGNATVYKQDQ
VAKKENGVTFKERGAWEPMVGRQENQLYHFANTYNTLVATVSDVGVKPEGTTIPLM

mRNA
CDS

mRNA

CDS

mRNA

CDS

mRNA

CDS

mRNA

CDS

mRNA

CDS

mRNA

GVRMSGGNPVLGLSYDSNNKCKWILLCCGGENKGLSIKQPKTHQVAIVLRNQTQ
GSAYVDGQVGGNGECALGEAAKQISHFYIGDGNADNKEVSVTVNVLVNRPLS
SAMTALNKKLSISKAKZKTEVTPQEVSEPATLETETPSSLGQGOQTPVLRITSE
NAGSGVLSYPAVSSVTTSPAVKESDQASGTSPEHSNVDDVDSSEGVQVDAETGD
AVCGSGTQQPSVGTATADTNASTAIEIMAHDDGTAVAPEVGAHSGEGTETEGTGQKR
EDIHAQDGEVKAALSSLGNSQNSSTSTVRGSGLLSLLLLLLGLLWGPAAAL"
/product="unknown"
/protein_id="AAP80762.1"
/db_xref="GI:32401137"
/translation="MAQORIOVTAGARGODKQVAKGFKRIRRGPNHPIVTRGHAS
TKCADSSGDRSKWSDDDATDERVSLPDSVGAIRYMKRLRAVSDAKQKLLPGGGVE
RSPFIALQMTVPVSPITPAKGLATMLAQGADMTAIQETWKPSEQIASMHTGEYVLHAQS
RIRKBCVAAVLVRKTLRSKRMPLITPQRDPSLDVVVVQVQFALQGNHDLIVANTYMRPP
QLTQPLGG"
/product="unknown"
/complement(<19732..>20760)
/product="unknown"
/protein_id="AAP80763.1"
/db_xref="GI:32401138"
/translation="WALKISAMSLREDTSWQTRRSASHARSSSGVRACTIFLFSLY
VHPLNLNLGSPGVTADMDHADDLPVINGHSREDTIPAAAMPFKLHARGOESGLVIN
PSKCEANWFTLPHTESNYDRKGKGLPVVAGCQVVPVVTMGASRTTKLLGMDLDPQLTL
NVAASQRISQPCIEHKEKGPDPRTFTIYGASKLRYSGLIWAIVATDSAKNEM
KQYATLARIIVSGVFTVPDESALLEANMPLHLVLCRLARLSIFENTRACQTDWRRRP
PPEPPRAGFRISPLRDELYAFVDATKYDQGITQSSPRENRPFRSSIPPMSAASAH
VTIGVELR"
/complement(<22600..>25278)
/product="unknown"
/complement(22600..25278)
/note="TcC2-ORF6; similar to Trypanosoma cruzi shed
acute-phase antigens deposited in PIR Accession Number
S23006 and encoded by GenBank Accession Number X57235;
also similar to Trypanosoma cruzi EST clones TENF0413 and
TENU5113 deposited in GenBank Accession Numbers AA556106
and AW330367"
/codon_start=1
/product="unknown"
/protein_id="AAP80764.1"
/db_xref="GI:32401139"
/translation="MGTVTVGASRMFLMFPVPLLLAICPSEPAHALAPGSSRVELPK
RQSKVPFEXDKGVTDRVHSPRLPALVNVGDVGMVAIADARYTSDNSLDTVAKYS
VDGEGWETQIALKNRSASSVSRVDPVTIVKGNKIYLVGTSNRSRYWTHGDARD
WDILLAVEGVTKTAGKITASIKWSPVSLKFKFPFAEMEGMHTNQFLGGAGVAIVAS
NGNLVYPQVTKNRKQVFSKIFYSEDDGKTKFGKGRSDFCGSEPTVALEWEGKLIINT
RVDNRKRLRWLRDLRLNLTDNORIYNVGQVSIKDENSASVSVLYKDDKLYCLHEINS
NEVSVLVARLVGELMIKSVLQSKWNDSHLSICTPADPAASSERGGCPAVTVG
LVGLTSQNSQNVWDAYANASTANARVNGLXFPAGVGGALAPVSCQGNQYR
FANHAETLVASVTIHEAPRASPLGASLDSGGKLLGLSYDERHQWQPTIYGSTPVT
PTGSETGKRYHVLLTWANKIGSVYIDGELLEGGQTVVPDQRTDISHFYVGGYGRS
DMPTIHSHTVNNVLLYNNRQNAEIKTLFLSQDLIGTEAMHDSRGDSAHGTPSTPD
SSAHGTPSTPDVDSAHGTPSTPDVDSAHGTPSTPDVDSAHGTPSTPDVDSAHGTPST
ADSSAHGTPSTPDVDSAHGTPSTPDVDSAHGTPSTPDVDSAHGTPSTPDVDSAHGTPS
TPAGSAHGTPSTPDVDSAHGTPSTPDVDSAHGTPSTPDVDSAHGTPSTPDVDSAHGTPS
VLTLPDGVALTSPSGGGGLLACALLLHLFPFNAVF"
/complement(<27726..>28667)

CDS	/product="unknown"			27726..28667					
	/notes="TcC2-ORF7; similar to Trypanosoma cruzi EST clones TENU3082, TENU0878, TENU0741, TENS2104, and TENU1990 deposited in GenBank Accession Numbers AI075490, AI667932, AI026529, AI562572, and AI057898"								
	/codon_start=1								
	/product="unknown"								
	/protein_id="AAP80765.1"								
	/db_xref="GI:32401140"								
	/translation="MAPARTHPSSNVTPSNOLCACHAHECVWDAFNVTADSGKKHTA GOROVKONTIILSLVHRITVDHAKSPSLGTHSPHTILQLEVVYASAAKTSRR ACQVKNQKASPTLLSRAPVGAERAGMGAPAAVEPFLVSLVSVLVLSSCVTRAAG GGVPLGVPSAAGSPGLAAPPPGDVGSAPPPAGVLRAVSDAASVCGAGERSPVA VGSPPAGDCRSDSPVSVKVDVSGFVFKLVGSFVAGPDPAGGGGGVAVVP FPADEFTFFVFSTPPGLPTLPCLSCWLSPPAGPPSLV"								
ORIGIN									
Query Match	90.0%; Score 1809.4; DB 3; Length 30095;								
Best Local Similarity	94.6%; Pred. No. 0;								
Matches 1873; Conservative	0; Mismatches 106; Indels 0; Gaps 0;								
Qy	22	CATCATCACAGCAGCGCTGGTGC	CGCGCGGCGAGCCATATG	GCACCCCGGATCGAGCGGA	81				
Db	25221	CTTCTTCTTGGCATCTGCC	CAGCGCCGCGACCGCTT	GGCACC	CGGATCGAGCGG	25162			
Qy	82	GTTGAGCTGTTTAAGCGGCAAG	CTCGAAGGTGCAATTTG	AAAGGCGGCGCAAGTCACC	141				
Db	25161	GTTGAGCTGTTTAAGCGGCAAG	CTCGAAGGTGCAATTTG	AAAGGCGGCGCAAGTCACC	25102				
Qy	142	GAGCGGTTGTCACTCGTTC	CGCTCCCGCCCTTGT	TAAATG	TGACGGGGTGATGGTT	201			
Db	25101	GATCGGGTTGTCACTCGTTC	CGCTCCCGCCCTTGT	TAAATG	TGACGGGGTGATGGTT	25042			
Qy	202	GCCATCGCGACGCTGCTAC	GAAACATCCAAATGACA	CACTCCCTCATTTGATAC	CGTGGCG	261			
Db	25041	GCCATCGCGACGCTGCTAC	GAAACATCCAAATGACA	CACTCCCTCATTTGATAC	CGTGGCG	24982			
Qy	262	AAGTACAGCGTGACGAT	GCGGAGAGCGTG	GGAGACCCAAATG	CCATCAAGAACAGTCGT	321			
Db	24981	AAGTACAGCGTGACGAT	GCGGAGAGCGTG	GGAGACCCAAATG	CCATCAAGAACAGTCGT	24922			
Qy	322	GCATCGTCTGTTTCTCGT	TGGTGATCCCA	CAGTGATTTG	TGAAGGCGAACAGCTTTAC	381			
Db	24921	GCATCGTCTGTTTCTCGT	TGGTGATCCCA	CAGTGATTTG	TGAAGGCGAACAGCTTTAC	24862			
Qy	382	GTCTGTGTGGAAGCTAC	ACAGTTTGAGAGCT	CTGACGCTCGCAT	GTGTGATGGAGA	441			
Db	24861	GTCTGTGTGGAAGCTAC	ACAGTTTGAGAGCT	CTGACGCTCGCAT	GTGTGATGGAGA	24802			
Qy	442	GACTGGGATATTCTG	CTTGCCTTGTGTGAG	GTCAAGTCC	ACTGCGGGCGGCAAGATA	501			
Db	24801	GACTGGGATATTCTG	CTTGCCTTGTGTGAG	GTCAAGTCC	ACTGCGGGCGGCAAGATA	24742			
Qy	502	ACTGCGAGTATCAAA	TGGGGAGCCCGTGT	CACTGAAGGAA	TTTTTCCCGGGGAAATG	561			
Db	24741	ACTGCGAGTATCAAA	TGGGGAGCCCGTGT	CACTGAAGGAA	TTTTTCCCGGGGAAATG	24682			
Qy	562	GAAGGAATGCA	CAAAATCAATTTCT	TGGCGGTG	CGAGTGTG	CGCATTTGGCGTCCAAC	621		
Db	24681	GAAGGAATGCA	CAAAATCAATTTCT	TGGCGGTG	CGAGTGTG	CGCATTTGGCGTCCAAC	24622		
Qy	622	GGGAATCTTG	TGTAACCTGTG	CAGGTTTAC	GAAACAA	AAAGAAAGCAAGTTTTTCC	681		
Db	24621	GGGAATCTTG	TGTAACCTGTG	CAGGTTTAC	GAAACAA	AAAGAAAGCAAGTTTTTCC	24562		
Qy	682	TTCTACTCGGAAG	ACGAGGCAAGACGT	GGAAGTTT	TGGGAGGGTAGGAGTGA	TTTTCG	741		
Db	24561	TTCTACTCGGAAG	ACGAGGCAAGACGT	GGAAGTTT	TGGGAGGGTAGGAGTGA	TTTTCG	24502		
Qy	742	TGCTCTGAAC	CTGTGGCCCTTG	ATGTTGGAGGGG	GAAGCTCATTA	AAAACTCGAGTTGAC	801		
Db	24501	TGCTCTGAAC	CTGTGGCCCTTG	ATGTTGGAGGGG	GAAGCTCATTA	AAAACTCGAGTTGAC	24442		

Qy	802	TATCGCGCGCTCTG	TGTGTACAGTCCAGT	GACATGGGGAA	TTTCGTGGGTGGAGGCTGTC	861	
Db	24441	CGGAAACGCGCGTCTG	TGTGTACAGTCCAGT	GACATGGGGAA	TATCGTGGGTGGAGGCTGTC	24382	
Qy	862	GGCAGCGCTCTCA	CGTGTGGGGCCCT	CACCAAAATCGAA	CCAGCCGGCAGTCAGAGC	921	
Db	24381	GGCAGCGCTCTCG	GTGTGGGGCCCT	CACCAAAATCGG	ACCAGCCGGCAGTCAGAGC	24322	
Qy	922	AGCTTCACTGCC	GTGACCATCGAG	GGAATTCGCTGT	TATGCTTTCACACACCCGCTGAAT	981	
Db	24321	AGCTTCACTGCC	GTGACCATCGAG	GGAATTCGCTGT	TATGCTTTCACACACCCGCTGAAT	24262	
Qy	982	TTTAAAGGGAAG	TGGCTGCGCGAC	CGACTGGAACCT	CTGCTGACGCGATAACAGCGGCAAT	1041	
Db	24261	TTTAAAGGGAAG	TGGCTGCGCGAC	CGACTGGAACCT	CTGCTGACGCGATAACAGCGGCAAT	24202	
Qy	1042	TATAAGCTTTGG	GCAAGTATCCAT	TGGTGTGATAA	AAATTCGCGCTACAGCTCCGCTCTGTAC	1101	
Db	24201	TATAAGCTTTGG	GCAAGTATCCAT	TGGTGTGATAA	AAATTCGCGCTACAGCTCCGCTCTGTAC	24142	
Qy	1102	AAGGATGATAAG	CTGTACTGTTTGC	ATGAGATCAACAG	TAAACGAGGTGTACAGCTTTGTT	1161	
Db	24141	AAGGATGATAAG	CTGTACTGTTTGC	ATGAGATCAACAG	TAAACGAGGTGTACAGCTTTGTT	24082	
Qy	1162	TTTTCGCGCTGT	TGGCGAGCTAC	CGEATCATTAA	ATCAGTGTCTGCTCAGTCTCTGGAAGAAT	1221	
Db	24081	TTTTCGCGCTGT	TGGCGAGCTAAT	GATTAAATCCG	TGCTGCTCAGTCTCTGGAAGAAT	24022	
Qy	1222	TGGGACAGC	CACCTGTTCAG	CATTTGCAC	CCCTGCTGATCCAGCGCTTCGTCGTGACAG	1281	
Db	24021	TGGGACAGC	CACCTGTTCAG	CATTTGCAC	CCCTGCTGATCCGCGCGCTTCGTCGTGACAG	23962	
Qy	1282	CGTGGTGTG	TCGCGCTGTCA	CACAGGTTGG	CTTTTGGCTTTTGTGCGACAGTGCC	1341	
Db	23961	CGTGGTGTG	TCGCGCTGTCA	CACAGGTTGG	CTTTTGGCTTTTGTGCGACAGTGCC	23902	
Qy	1342	ACCAAAACG	GAATGGGAGG	ATCGTACCG	CTGCTGCTCAAACGCAACGCGCAAAATTCGCGAG	1401	
Db	23901	TCCAAAAAG	TATGGGAGG	ATCGTACCG	CTGCTGCTCAAACGCAACGCGCAAAATTCGCGAG	23842	
Qy	1402	AGGGTTCG	AAACGGTTTGA	AGTTTGGGGGG	TTTGGGAGGGGCGCTTTTGCGCGGTGAGC	1461	
Db	23841	AGGGTTCG	AAACGGTTTGA	AGTTTGGGGGG	TTTGGGAGGGGCGCTTTTGCGCGGTGAGC	23782	
Qy	1462	CACGAGGGC	AGAAATCAAC	CGGTATCACTT	TGCAAAACACGCGCTTCA	CGCTGGTGGCGTCG	1521
Db	23781	CACGAGGGC	AGAAATCAAC	CGGTATCGT	TTTGCAAAACACGCGCTTCA	CGCTGGTGGCGTCG	23722
Qy	1522	GTGCAAT	CAACGAGGT	TCGAGCTGCG	AGTCTCTTTGCTGGGTGCGAGCTTGGA	CTCT	1581
Db	23721	GTGCAAT	CAACGAGGT	TCGAGCTGCG	AGTCTCTTTGCTGGGTGCGAGCTTGGA	CTCT	23662
Qy	1582	TCCTGTGGC	AAAAAATCTCT	CTGGGGCTCTG	TACGACGAGGAAGCAC	CAGTCGGCAGCCAATA	1641
Db	23661	TCCTGTGGC	AAAAAATCTCT	CTGGGGCTCTG	TACGACGAGGAAGCAC	CAGTCGGCAGCCAATA	23602
Qy	1642	TACGGAT	CAACGCGGT	GACGCGGAT	CGTGGGAGATGGGTAA	GAGGTACCA	1701
Db	23601	TACGGAT	CAACGCGGT	GACGCGGAT	CGTGGGAGATGGGTAA	GAGGTACCA	23542
Qy	1702	GTTCTTAC	GATGCGGAATA	AAATTTGGT	TTCGGTGTACAT	TGATGGAGAACCTCTG	1761
Db	23541	GTTCTTAC	GATGCGGAATA	AAATTTGGT	TTCGGTGTACAT	TGATGGAGAACCTCTG	23482
Qy	1762	TGAGG	CAGACCGGT	GTGCGAG	CGCTGACATCTCC	CACTTCTAGTTGGC	1821
Db	23481	TGAGG	CAGACCGGT	GTGCGAG	CGCTGACATCTCC	CACTTCTAGTTGGC	23422
Qy	1822	GGGTAT	GGAAGGAT	GATATGCC	CAACATTAAGCC	AGTGA	1881
Db	23421	GGGTAT	GGAAGGAT	GATATGCC	CAACATTAAGCC	AGTGA	23362

gene /lab_host="Vero cells"
133..139
RBS /gene="trans-sialidase"
133..139
/gene="trans-sialidase"
/note="Kozak sequence ACCATGG at the position indicated"
/citations[2]

ORIGIN

Query Match 87.3%; Score 1753.8; DB 3; Length 2133;
Best Local Similarity 94.3%; Pred. No. 0;
Matches 1831; Conservative 0; Mismatches 107; Indels 3; Gaps 1;

QY 22 CATCATCACAGAGCGGCTGTGGCCGCGGAGCCATATGGCACCCCGATCGAGCCGA 81
DB 193 CTTCTCTTCGCGATCGCCCGAGCGAGCGCGGTACGCCCTTGGCACCGGATCGAGCCGA 252
QY 82 GTTGAGCTGTTTAAGCGGCAAGCTCGAAGTGTGCATTGAA---AAGGCGGCAAGTC 138
DB 253 GTTGAGCTGTTTAAGCGTAAGAAATTCGACGGTGCCTGTTGAAGCAAGGCCGCAAGTC 312
QY 139 ACCGAGCGGCTGTCCACTCGTTCCGCTCCCGCCCTCTGTTAATGTGACGCGGTGATG 198
DB 313 ACCGAGCGGCTGTCCACTCGTTCCGCTCCCGCCCTCTGTTAATGTGACGCGGTGATG 372
QY 199 GTTGCCATCGCGAGCGCTCGTACGAAACHATCCAATGACAACTCCCTCATTTGATACGGTG 258
DB 373 GTTGCCATCGCGAGCGCTCGTACGACACATCCAATGACAACTCCCTCATTTGATACGGTG 432
QY 259 GCGAAGTACAGCTGACGATGGGAGACGTGGAGACCCAAATGGCCATCAAGAACAGT 318
DB 433 GCGAAGTACAGCTGACGATGGGAGACGTGGAGACCCAAATGGCCATCAAGAACAGC 492
QY 319 CGTGCACTCGTCTGTTCTCGTGTGGTGGATCCACACAGTGTGTAAGGGCAACAAGCTT 378
DB 493 CGTGATCGTCTGTTCTCGTGTGGTGGATCCACACAGTGTGTAAGGGCAACAAGCTT 552
QY 379 TACGTCTGTGGAGCTACAAAGTTTCAGAGAGCTACTGAGCGTGCATGTGATGTCG 438
DB 553 TACGTCTGTGGAGCTACTAGTTTCAGAGAGCTACTGAGCGTGCATGTGATGTCG 612
QY 439 AGAGACTGGGATATCTGCTGCGGTGGTGGAGTCAAGAGTCCACTGCGGCGGCAAG 498
DB 613 AGAGACTGGGATATCTGCTGCGGTGGTGGAGTCAAGAGTCCACTGCGGCGGCAAG 672
QY 499 ATAACTGCGAGTATCAAAATGGGAGCCCGCTGTCTACTGAAGGAAATTTTCCCGGGGAA 558
DB 673 ATAACTGCGAGTATCAAAATGGGAGCCCGCTGTCTACTGAAGGAAATTTTCCCGGGGAA 732
QY 559 ATGAAGAGGATGACACAAATCAATTTCTTGGCGGTGCGAGGTGTGCAATTTGGCGGTCC 618
DB 733 ATGNAAGGATGACACAAATCAATTTCTTGGCGCGCGGTGTGCAATTTGAGCGTCC 792
QY 619 AACGGAAATCTGTGTACCTGTGCGAGTTTACGAAACAAAGAGCAAGTTTTCCTCAAG 678
DB 793 AACGGAAATCTGTGTACCTGTGCGAGTTTACGAAACAAAGAGCAAGTTTTCCTCAAG 852
QY 679 ATCTTCTACTCGGAAGACGAGGCAAGAGCTGGAAGTTTGGGAGGAGGTAGAGTATTTT 738
DB 853 ATCTTCTACTCGGAAGATGATGCAAGACGTGGAAGTTTGGGAGGAGGTAGAGTATTTT 912
QY 739 GGCTGTCTGAACTGTGGCCCTTGTAGTGGGAGGGGAAAGCTCATATAAACACTCGAGTT 798
DB 913 GGCTGTCTGAACTGTGGCCCTTGTAGTGGGAGGGGAAAGCTCATATAAACACTCGAGTT 972
QY 799 GACTATCGCGCGCTGTGTGTACGAGTCAAGTGAATGGGGAATTCGTGGGTGAGGCT 858
DB 973 GACTGGAACCGCGCTGTGTGTACGAGTCAAGTGAATGGGGAATTCGTGGGTGAGGCT 1032
QY 859 GTCGGCAGCTCTCACTGTGTGGGCGCCCTCACCAAAATCGAAACAGCCCGCAGTCA 918
DB 1033 GTCGGAACCGCTCTCGCGTGTGTGGGCGCCCTCACCAAAATCGAAACAGCCCGCAGTCA 1092

QY 919 AGCAGCTTCACTGCGCGTGACCATCGAGGAATCGGTGTATGCTCTTTCACACACCGCGTG 978
DB 1093 AGCAGCTTCACTGCGCGTGACCATCGAAGGAATCGGTGTATGCTCTTTCACACACCGCGTG 1152
QY 979 AATTTTAAAGGAAGGTGGCTGCGCGACCGACTCAACTCTGGCTGACGGAATAACACAGCGC 1038
DB 1153 AATTTTAAAGGAAGGTGGCTGCGCGACCGACTCAACTCTGGCTGACGGAATAACACAGCGC 1212
QY 1039 AATTTATAACGTTGGGCAAGTATCCATTTGGTGTATGAAATAATTCGCGCTACAGCTCGCTCGT 1098
DB 1213 AATTTATAACGTTGGGCAAGTATCCATTTGGTGTATGAAATAATTCGCGCTACAGCTCGCTCGT 1272
QY 1099 TACAAGGATGATTAAGCTGTACTGTTGTCATGAGATCAACAGTAACGAGGTGTACAGCTT 1158
DB 1273 TACAAGGATGATTAAGCTGTACTGTTGTCATGAGATCAACACGAGGAGGTGTACAGCTT 1332
QY 1159 GTTTTTCGCGCGCTGTTGGCGAGCTACGATCATTAATAATCAGTGTGCGAGTCTCTGGAAG 1218
DB 1333 GTTTTTCGCGCGCTGTTGGCGAGCTACGATCATTAATAATCAGTGTGCGAGTCTCTGGAAG 1392
QY 1219 AATTTGGGACAGCCACCTGTCCAGCATTTTGACCCCTGCTGATCCAGCGCTTTCGTGCTCA 1278
DB 1393 AATTTGGGACAGCCACCTGTCCAGCATTTTGACCCCTGCTGATCCAGCGCTTTCGTGCTCA 1452
QY 1279 GAGCGTGTGTTGGTCCCGCTGTCAACAAGTTGGTGGTCTTGTGGCTTTTGTGCGCAAGT 1338
DB 1453 GAGAGTGTGTTGGTCCCGCTGTCAACAAGTTGGTGGTCTTGTGGCTTTTGTGCGCAAGT 1512
QY 1339 GCCACCAAAACCGAATGGGAGGATGCTACCGCTGCTCAACGACGAGCAAAATGCG 1398
DB 1513 GCCTCCAAAACGATATGGGAGGATGCTACCGCTGCTCAACGAGCAAGCAAAATGCG 1572
QY 1399 GAGAGGTTCCGAAACGTTTGAAGTTTGGCGGGGTTGGCGAGGGCGCTTTTGCCCGGTG 1458
DB 1573 GAGAGGTTCCGAAACGTTTGAAGTTTGGCGGGGTTGGCGAGGAGCGCTTTTGCCCGGTG 1632
QY 1459 AGCAGCAGGGGAGAGATCAACGCTATCACTTTGCAAAACACGCGTTTCAAGCTGGTGCG 1518
DB 1633 AGCAGCAGGGGAGAGATCAAGCGGTATCGTTTTTGCAAACACACGCTTTCAGCTGGTGCG 1692
QY 1519 TCAGTACCATTCACAGAGTTCCGAGCTCGCGAGTCTTTGCTGGGTGCGAGCTCGAC 1578
DB 1693 TCAGTACCATTCACAGAGTTCCGAGGCTCCGAGGCGCGAGTCCCTTGTGGGTGCGAGCTCGAC 1752
QY 1579 TCTTCTGTGGGCAAAAACCTCTCTGGGCTCTCGTACGAGAGAGCAACAGTGGCAGCCA 1638
DB 1753 TCTTCTGTGGGCAAAAACCTCTCTGGGCTCTCGTACGAGAGAGCAACAGTGGCAGCCA 1812
QY 1639 ATATACGATCAACCGCGGTGACCGGATCGTGGAGATGGGTAAAGGTATCCAC 1698
DB 1813 ATATACGATCAACCGCGGTGACCGGATCGTGGAGAGCGGTAAAGGTATCCAC 1872
QY 1699 GTGGTTCTTACGATGGGAATAAATTTGGTTCGCTGTACATTCATTCGAGAACCTCTGGAG 1758
DB 1873 TTGGTTCTTACGATGGGAATAAATTTGGTTCGCTGTACATTCATTCGAGAACCTCTGGAG 1932
QY 1759 GGTTCAGGCGACAGCCCTTGTGCGAGACGGGAGACGCTGACATCTCCCACTTCTAGCTT 1818
DB 1933 GGTTCAGGCGACAGCCCTTGTGCGAGACGGGAGACGCTGACATCTCCCACTTCTAGCTT 1992
QY 1819 GCGCGGTATGGAAGGATGATATGCCAACCAATAAGCCACGTGACGCTGAATAATGTTCTT 1878
DB 1993 GCGCGGTATGGAAGGATGATATGCCAACCAATAAGCCACGTGACGCTGAATAATGTTCTT 2052
QY 1879 CTTTACAAACGCTCAGCTGATGCGGAGGATCAGGACCTTGTCTTGTGAGCAGGACCTG 1938
DB 2053 CTTTACAAACGCTCAGCTGATACCGAGGAGATCAGGACCTTGTCTTGTGAGCAGGACCTG 2112
QY 1939 ATTGGCACCGAAGCACACATG 1959
DB 2113 ATTGGCACCGAAGCACACATG 2133

RESULT 13
TRBNCNA
LOCUS TRBNCNA 5403 bp DNA linear INV 11-MAR-1994
DEFINITION T. cruzi neuraminidase (TCNA) gene, complete cds.
ACCESSION M61732
VERSION M61732.1 GI:162302
KEYWORDS neuraminidase.
SOURCE Trypanosoma cruzi
ORGANISM Trypanosoma cruzi
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
REFERENCE 1 (bases 1 to 5403)
AUTHORS Pereira,M.E., Mejia,J.S., Ortega-Barria,E., Matzilevich,D. and Prioli,R.P.
TITLE The Trypanosoma cruzi neuraminidase contains sequences similar to bacterial neuraminidases, YWD repeats of the low density lipoprotein receptor, and type III modules of fibronectin
J. Exp. Med. 174 (1), 179-191 (1991)
91277609
1711561
COMMENT Original source text: T. cruzi (strain Silvio) DNA, clone 7P.
FEATURES
Location/Qualifiers
1..5403
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/db_xref="taxon:5693"
484..3972
/gene="TCNA"
484..3972
/gene="TCNA"
/EC_numbers="3.2.1.18"
/codon_start=1
/product="neuraminidase"
/protein_id="AAA30255.1"
/db_xref="GI:162303"
/translation="MVALADARYETSSNSLIDTVAKYSVDDGETWETQIAIKNSRVS
SVSVWPTVIVKGNKLYLVGSYISRSYSSWGGDARDWDILLAYGEVTKSTAGGKI
TASIKWSPVSLKFFFAEMEGMHTNQFLGAGVAIVASGNLVIPVQVTKRQKQVFS
KIFYSDGDKTKFKGSDPFCSEPEVALEWEGKLIINTVDWKRLVTVESSDMEKPF
VEAVGTIRYVWGPSPKSPQSGSOTFAVTEGMVLMFLPHPLNFKRCVRLRLNLWL
TDNORIYVNGVSGIDNSAVSVLYKDDKLYCLHEINTDEVYSLVPLRVGLRLIK
SVLSRWKNTATCAPALLOPLRQRVVPVPLSRLLVLLAFRCRLRMGGGYSYRC
VNASTAERVRNGKLPAGVGGALMPVQSQNRFRANFANHTLVASVTHIEAPRA
ASPLGASLSSGKLLGLSYDEKQWQPIYGSTPTVPTGSMETQKRYHLVLTWANK
IGSVIIDEELLESQVDPDGRTPDISHFYGVGYKRSMDPTISHVTNNVLLYNRRO
LNTBEIRTLFLQDLGTEAHMDSDDSSAHSPTSPADSSAHSPTSPVDSAHSTP
STPADSSAHSPTSPVDSAHSTPSPADSSAHSPTSPADSSAHSPTSPVDSAHSTP
TPSTPVDSSAHGAPSPADSSAHSPTSPVDSAHSTPSPADSSAHSPTSPVDSAHSTP
HSTPSTPADSSAHSPTSPVDSAHSTPSPADSSAHSPTSPADSSAHSPTSPVDS
SAHSTPSTPVDSSAHSTPSPVDSAHSTPSPADSSAHSPTSPADSSAHSPTSPVDS
DSSAHSPTSPADSSAHSPTSPVDSAHSTPSPADSSAHSPTSPADSSAHSPTSP
PADSSAHSPTSPVDSAHSTPSPADSSAHSPTSPADSSAHSPTSPADSSAHSPT
STPVDSSAHSPTSPADSSAHSPTSPADSSAHSPTSPADSSAHSPTSPADSSAHS
TPSTPADSSAHSPTSPADSSAHSPTSPVDSAHSTPSPADSSAHSPTSPADSSA
HSTPSTPADSSAHSPTSPADSSAHSPTSPADSSAHSPTSPADSSAHSPTSPVDS
LCACALLHVFVFAV"

Query Match 85.0%; Score 1708.2; DB 3; Length 5403;
Best Local Similarity 93.3%; Pred. No. 0;
Matches 1852; Conservative 0; Mismatches 123; Indels 10; Gaps 6;
QY 22 CATCATCAGCAGCGCGCTGGTCCGCGCGGAGCCATATGCGACCCGGATCGAGCCGA 81
DB 308 CTTCCTCTTCGATCTGCCCGCAGCGAGCCCGGTAGCGCCCTGGCAGCCGGATCGAGCCGA 367
QY 82 GTTGAGCTGTTTAAGCGGCAAGCTCGAAGGTGCCATTGAA---AAGCGGCGCAAGTC 138
DB 368 GTTGAG-GGTTTAAGGTAAGAATTCGACGGTGGCGTTTGAACAAGCGCGCAAGTC 426
QY 139 ACGAGCGGTTGTCACCTCGTTCGCTCCCGCTCCCGCTTGTAAATGTGGAAGCGGCTGATG 198
DB 427 ACGAGCGGTTGTCACCTCGTTCGCTCCCGCTTGTAAATGTGGAAGCGGCTGATG 486

QY 199 GTTGCCATCGGAGCGCTCGCTACGAAACATCCAAATGACAACTCCCTCATTTGATCGGTG 258
DB 487 GTTGCCATCGGAGCGCTCGCTACGAAACATCCAGTGAAACCTCCCTCATTTGATCGGTG 546
QY 259 GCGAAGTACAGCGTGGACGATGGGGAGACGTTGGGAGACCCCAATTCGCCATCAAGAACAGT 318
DB 547 GCGAAGTACAGCGTGGACGATGGGGAGACGTTGGGAGACCCCAATTCGCCATCAAGAACAGC 606
QY 319 CGTGCACTCGCT 378
DB 607 CGTGCACTCGCT 666
QY 379 TACGTCTCTGTGGAAAGCTACAACTGCGAGAGCTACTGCGAGAGCTACTGCGAGCTGCGATGATGCG 438
DB 667 TACGTCTCTGTGGAAAGCTACTATAGTTTCGAGAGCTACTGCGTCTGCGATGATGATGCG 726
QY 439 AGAGACTGGGATATTTCTGCTTGGCTTGGTGGAGTTCAGAACTCCACTGCGGGCGGCAAG 498
DB 727 AGAGACTGGGATATTTCTGCTTGGCTTGGTGGAGTTCAGAACTCCACTGCGGGCGGCAAG 786
QY 499 ATAACCTGCGAGTATCAAAATGGGGAGCGCCCGTGTCTCACTGAAGGAATTTTCCCGCGGNA 558
DB 787 ATAACCTGCGAGTATCAAAATGGGGAGCGCCCGTGTCTCACTGAAGGAATTTTCCCGCGGNA 846
QY 559 ATGGAAGGAATGCACACAAATCAATTTCTTGGCGGTGCGAGTGTGTCATTTGTGCGCGTCC 618
DB 847 ATGGAAGGATGCACACAAATCAATTTCTTGGCGCGCGGTGTTGTCATTTGTGCGCGTCC 906
QY 619 AACGGGATCTTGTGTACCCCTGTGCGAGGTACGACAAAGAGCAAGTATTTTCCCGCGGNA 678
DB 907 AACGGGATCTTGTGTACCCCTGTGCGAGGTACGACAAAGAGCAAGTATTTTCCCGCGGNA 966
QY 679 ATCTTCTACTCGGAAGAGTGGGAGCGTGGAGTGTGCGAGGTGAGGAGTATTTT 738
DB 967 ATCTTCTACTCGGAAGAGTGGGAGCGTGGAGTGTGCGAGGTGAGGAGTATTTT 1086
QY 739 GCGTCTCTGACCTGTGCGCTTGTGAGTGGGAGGAGGAGTCTATCAATAACATCGAGTT 798
DB 1027 GCGTCTCTGAACTGTGCGCTTGTGAGTGGGAGGAGGAGTCTATCAATAACATCGAGTT 1086
QY 799 GACTATCCCGCGCTGTGTGTAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 858
DB 1087 GACTGTGAAACCGCTGT 1146
QY 859 GTGCGACCGCTCTCACTGT 918
DB 1147 GTGCGAACCGCTCTCGCTGT 1206
QY 919 AGCAGCTTCACTGCGCGT 978
DB 1207 AGCAGCTTCACTGCGCGT 1266
QY 979 AATTTTAAGGAGAGT 1038
DB 1267 AATTTTAAGGAGAGT 1326
QY 1039 ATTTTATAACCTGTGGGCAAGTATCCATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1098
DB 1327 ATTTTATAACCTGTGGGCAAGTATCCATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1386
QY 1099 TACAAGGATGATAAGCTGTACTGTTTGTGATGATCAACAGTAAACAGGTTGTACAGCCTT 1158
DB 1387 TACAAGGATGATAAGCTGTACTGTTTGTGATGATCAACAGGATCAACAGGATGATGATGATGAT 1446
QY 1159 GTTTTTCGCGCTGT 1218
DB 1447 GTTTTTCGCGCTGT 1506
QY 1219 AATTTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1278
DB 1507 AATTT-GGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1565

Qy 1279 GAGCGTGGTGTGCTCCCGTGTGTCACCAAGGTTGGTCTTGTGTGGCTTTTGTGTGCGCAGT 1338
Db |||||
1566 GAGAGTGGTGTGCTCCCGTGTGTCACCAAGGTTGGTCTTGTGTGGCTTTTGTGTGCGCAGT 1624
Qy |||||
1339 GCCACCAAAACCGAATGGGAGGATGCTACCGCTGCGTCAACGCAAGCAGCGCAAAATGCG 1398
Db |||||
1625 GCCTCCCAAAACGATGGGAGGAT-CGTACCGCTGCGTCAACGCAAGCAGCGCAAAATGCG 1683
Qy 1399 GAGAGGTTCCGAACCGTGTGCAAGTGTGCGGGGTGGCGAGGGCGCTTTTGGCCGGTG 1458
Db |||||
1684 GAGAGGTTCCGAACCGTGTGCAAGTGTGCGGGGTGGCGAGGGAGCGCTTTGGCCGGTG 1743
Qy 1459 AGCAGAGGGGCGAATCAACGATATCACTTTGCAAAACCAACGCTTCAAGCTGTGGCG 1518
Db |||||
1744 AGCAGAGGGGCGAATCAACGATATCACTTTGCAAAACCAACGCTTCAAGCTGTGGCG 1803
Qy 1519 TCGGTACGATTCACAGAGTTCGAGCGTCCGAGTCTTGTGCGGTGGAGCTGCGAC 1578
Db |||||
1804 TCGGTACGATTCACAGAGTTCGAGCGTCCGAGGCGCGAGTCCCTTGTGCGGTGGAGCTGCGAC 1863
Qy 1579 TCTTCTGGTGGCAAAACCTCCTGGGCTCTCGTACGAGGAGCAACCAAGTGGCAGCCA 1638
Db |||||
1864 TCTTCTGGGCGCAAAACCTCCTGGGCTCTCGTACGAGGAGCAACCAAGTGGCAGCCA 1923
Qy 1639 ATATACGATCAACCGCGTGAAGCCGACCGGATCGTGGAGATGGGTAAGAGGTACAC 1698
Db |||||
1924 ATATACGATCAACCGCGTGAAGCCGACCGGATCGTGGAGAGCGGTAAGAGGTACAC 1983
Qy 1699 GTGGTCTTACGATGGCAATAAATGGTTCGGGTACATTCGATGAGAACCTTCTAGCTT 1758
Db |||||
1984 TTGGTCTTACGATGGCAATAAATGGTTCGGGTACATTCGATGAGAACCTTCTAGCTT 2043
Qy 1759 GGTTCAGGCGAGACCGTGTGTCAGAGCGGAGGACGCTGACATCTCCCACTTCTAGCTT 1818
Db |||||
2044 GGTTCAGGAGACCGTGTGTCAGAGCGGAGGAGCGCTGACATCTCCCACTTCTAGCTT 2103
Qy 1819 GCGCGGTATGGAAGGATGATATGCCAAACATAGCCACGTCGAGCGTGAATATGTTCTT 1878
Db |||||
2104 GCGCGGTATGGAAGGATGATATGCCAAACATAGCCACGTCGAGCGTGAATATGTTCTT 2163
Qy 1879 CTTTACAAAC--CGTACGCTGAATGCGAGGAGATCAGGACCTGTTCTTGGCCAGGAC 1935
Db |||||
2164 CTTTACAAACGAGACGACTGAATACCGAGGAGATCAGGACCTGTTCTTGGCCAGGAC 2223
Qy 1936 CTGATTCGCAAGGACACATGCGGAGCAGCAGCGGAGGAGTGAAGAAAGTACGCC 1995
Db |||||
2224 CTTATTTGGCAGGAGACACATGGAAGCAGCAGCAGCAGCAGTGGCCACAGTACGCC 2283
Qy 1996 GGATC 2000
Db |||||
2284 TCAAC 2288

RESULT 14
D50686
LOCUS
DEFINITION Trypanosoma cruzi TCTS-193 gene for trans-sialidase homolog,
partial cds.
ACCESSION D50686
VERSION D50686.1 GI:840709
KEYWORDS
SOURCE Trypanosoma cruzi
ORGANISM Trypanosoma; Kinetoplastida; Trypanosomatidae;
Eukaryota; Euglenozoa; Schizotrypanum.
REFERENCE
1 Smith, L.E., Uemura, H. and Eichinger, D.
Isolation and expression of an open reading frame encoding
sialidase from Trypanosoma rangeli
Mol. Biochem. Parasitol. 79 (1), 21-33 (1996)
JOURNAL 97001677
MEDLINE PUBMED 894669
REFERENCE 2 (bases 1 to 2100)

AUTHORS
TITLE
JOURNAL

Uemura, H.
Direct Submission
Submitted (27-MAY-1995) Haruki Uemura, Nagasaki University,
Institute of Tropical Medicine; 1-12-4 Sakamoto, Nagasaki, Nagasaki
852, Japan [E-mail: H-Uemura@cc.nagasaki-u.ac.jp,
Tel: 81-958-49-7837, Fax: 81-958-49-7805]

FEATURES

source

1. .2100
/organism="Trypanosoma cruzi"
/mol_type="genomic DNA"
/strain="ny"
/db_xref="taxon:5693"
/clone="193"
1. .2100
/gene="TCTS-193"
1. .>2100
/genes="TCTS-193"
/function="one of the trans-sialidase gene family"
/codon_start=1
/product="trans-sialidase homolog"
/protein_id="BAA09335.1"
/db_xref="GI:840710"
translation="MPRLLLYLVLPPPLNFCNPVPAHALAPSSRIKLPKRNSTVP
PEPKGKTERVHVSFRLPALVDVGVMAIADARYDTSDNLSLIDTVKLVSDGETW
ETQIGIKNSVSVSVVDPTVIVKGNKLYLVGVSYNSRSYTWQPDGSDMRPLAV
GVYTNKKELFSKIFYSEDDGKTWNFSEGRDFCEPVLWEWENKFIINRVDRR
LVYSGDMKEPWEAVGTLRSVWGSPSKSNQGSQSFYATVTEGRVMLFTHPLNFK
GWLRLHLWLTNORIVNVGVSGIDENSAYSSVLYKDDKLYCLHIEITNDEVYSIV
PARLVGELRLIKSVRSWKWDHSLISICTPADPAASSRSGGCPAVTTVGLAGFLSH
SATQWEDAIRCNASTNAEVRPNFAGVGGALWFPVQCGQNRTHPANYEFT
LVASVTIHEAPRNATPLLGSALSDSGKKLLSYDEKHQWEPYIGSTPTVPGSWET
GKYHVALLWNRNGSVYIDGVLKSGGTVPDERTPDISHFYIGSYNSNMPTISH
LTANNVLLNROLNABEIRTLFLSQDLIGTEGILDSDDGDKEDKSDSEYDDGAHSNS
SFFGGDASSVSAGRNLP"

gene

CDS

ORIGIN

Query Match 82.9%; Score 1666; DB 3; Length 2100;
Best Local Similarity 91.7%; Pred. No. 0;
Matches 1762; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

Qy 62 TGCCACCGGATCGAGCGGAGTTGAGCTGTTTAAGCGGCAAGCTCGAAGTGCCATTG 121
Db |||||
77 TGCCACCGGATCGAGTCGAATGAACTGTTTAAGCGTAAGAAATTCACGGTGCCGTTG 136
Qy 122 AAAAGGCGGCAAGTCACCGAGCGGTTGTCACTCGTTCCGCTCCCGCCCTCTGTTA 181
Db |||||
137 AAGCAAGGCAAGTCACCGAGCGGTTGTTCACTGTTCCGCTCCCGCCCTCTGTTG 196
Qy 182 ATGTGACGGGTGATGGTTGCCATCGCGACGCTCGCTACGAAACATCCAATGACAAT 241
Db |||||
197 ATGTGGATGGGTGATGGTTGCCATTCGACAGCGCCGCTACGACATCCAATGACAAT 256
Qy 242 CCCTCATTTGATCGTGGCGAAGTACAGCTGACGATGGGAGAGCTGGGAGACCCAAA 301
Db |||||
257 CCCTCATTTGATCGTGGTAGTGAAGTACAGCGTGGAGAGCTGGGAGACTCAA 316
Qy 302 TTCCCATCAAGAACAGTCGTCATCGTCTGTTCTCGTGTGGTGAATCCACACAGTGA 361
Db |||||
317 TTGGCATCAAGAACAGTCGTCATCGTCTGTTCTCGTGTGGTGAATCCACACAGTGA 376
Qy 362 TGAAGGCGCAACAGCTTTACGTCCTGTTGGAAGTCAACAGTTGAGAGGCTACTGGA 421
Db |||||
377 TGAAGGCGCAACAGCTTTACGTCCTGTTGGAAGTCAACATAGTTTCGAGGAGCTACTGGA 436
Qy 422 CGTGCAGTGGTATGCGAGAGACTGGGATATTCTGTTGCCCTGCTGGTGGAGTCA 481
Db |||||
437 CTTGGCAGCTGATGGAAGCGACTGGAGACTATGCTTGGCTTGGTGGAGTTACGAAGT 496
Qy 482 CCACTCGGCGGCAAGATAACTCGCAGTATCAATGCGGGAGCCCGTGTCTACTGAAG 541
Db |||||
497 CCACTCGACGCGCAAGACAACCTGCGAATATCATGGGGAGCCCGCTGTCTACTGAAG 556


```
QY 542 AATTTTCCCGCGGAAATGGAAGAAATGCACAAATCAATTTCTTGGCGGTGCAGGTG 601
Db 557 AATTTTTCGCGCAGAAATGGAAGAAATGCACAAAGCAATTTCTCGGCGGTGCAGGTG 616
QY 602 TTGCAATTTGGCGGTCAACCGGAAATCTTGTGTACCTGTGCAGGTTCAGAAACAAAGA 661
Db 617 TTGCAATTTGGCGGTCAACCGGAAATCTTGTGTCCCTGTGCAGGTTCAGAAACATGAAA 676
QY 662 AGCAAGTTTTTTCAGAGATCTTCTACTCGGAAGACGAGGSCAAGAGCTGGAAGTTTGGG 721
Db 677 AAGAACTTTTCTCCAAGATCTTCTACTCGGAAGATGATGCAAGAGCTGGAATTTCTCAG 736
QY 722 AGGTAGGAGTATTTTGGCTGTCTGAACTCTGAACTGTGCGCCCTTGAGTGGAGGGAAGCTCA 781
Db 737 AGGTAGGAGGAGTATTTTGGCTGTCTGAACTGTGCGCCCTTGAGTGGAGGGAAGCTCA 796
QY 782 TCATAAACACTCCAGTTGACTATCGCGCGCTGTGTGTACAGTGTGATGATGATGATGATG 841
Db 797 TCATAAACACTCCAGTTGACTATCGCGCGCTGTGTGTACAGTGTGATGATGATGATGATG 856
QY 842 ATTCTGGGTGGAGGTGTTCGACAGCTCTCACTGTGTGCGGCGCCCTCACCAAAATCGA 901
Db 857 AACCGTGGGTGGAGGTGTTCGACAGCTCTCACTGTGTGCGGCGCCCTCACCAAAATCGA 916
QY 902 ACCAGCCCGGAGTCAGAGAGCTTCACTGCGTGTGACCAATCGAGGGAATGCGTTATGC 961
Db 917 ACCAGCCCGGAGTCAGAGAGCTTCACTGCGTGTGACCAATCGAGGGAATGCGTTATGC 976
QY 962 TCTTACACACCCGCTGAAATTTTAAAGGAAGTGTGCGGAGCGCTGCACTGAACTCTGGC 1021
Db 977 TCTTACACACCCGCTGAAATTTTAAAGGAAGTGTGCGGAGCGCTGCACTGAACTCTGGC 1036
QY 1022 TGACGGATAACAGCGCATTTTAAAGTGGCAAGTATCACTGTTGATGATGAAATTCG 1081
Db 1037 TAAAGGATAACAGCGCATTTTAAAGTGGCAAGTATCACTGTTGATGATGAAATTCG 1096
QY 1082 CCTACAGCTCCGCTGTGACAGGATGATAGCTGTACTGTTGATGATGATGATGATGATG 1141
Db 1097 CCTACAGCTCCGCTGTGACAGGATGATAGCTGTACTGTTGATGATGATGATGATGATG 1156
QY 1142 ACGAGGTGTACAGCTTGTGTTTGGCGCCCTGTTGCGAGTACGATCATTAATTCAG 1201
Db 1157 ACGAGGTGTACAGCTTGTGTTTGGCGCCCTGTTGCGAGTACGATCATTAATTCAG 1216
QY 1202 TGCTCACTCTCTGGAAGATTTGGACAGCCACTGTCCAGATTTTGACCCCTGCTGATC 1261
Db 1217 TGGTGGCTTCTGGAAGATTTGGACAGCCACTGTCCAGATTTTGACCCCTGCTGATC 1276
QY 1262 CAGCGCTTCTGTGACAGCGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1321
Db 1277 CTGCGCTTCTGTGACAGCGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1336
QY 1332 GCTTTTGTGACAGTCCCAACCAACCGAATGGAGGATGCGTACCGTGGTCAACG 1381
Db 1337 GCTTTTGTGACAGTCCCAACCAACCGAATGGAGGATGCGTACCGTGGTCAACG 1396
QY 1382 CAAGACGGCAAAATCGGAGAGGTTCGGAACGGTTTGAAGTTTGGGGGTTGGCGAG 1441
Db 1397 CAAGACGGCAAAATCGGAGAGGTTCGGAACGGTTTGAAGTTTGGGGGTTGGCGAG 1456
QY 1442 GGGCGCTTTGGCGGTGAGCCAGCGGGCAGAAATCAACGGTATCACTTTGCAAAACGAG 1501
Db 1457 GGGCGCTTTGGCGGTGAGCCAGCGGGCAGAAATCAACGGTATCACTTTGCAAAACGAG 1516
QY 1502 CGTTACGCTGTGGGTGCGTACGATTTACGAGGTTCCGAGGTTCCGAGTCCGAGTCTTTGC 1561
Db 1517 AGTTACGCTGTGGGTGCGTACGATTTACGAGGTTCCGAGGTTCCGAGTCTTTGC 1576
QY 1562 TGGGTGCGAGCTGGAATCTTCTGTGTGGCAAAACTCTCTGGGCTCTCGTACGACGAGA 1621
Db 1577 TGGGTGCGAGCTGGAATCTTCTGTGTGGCAAAACTCTCTGGGCTCTCGTACGACGAGA 1636
QY 1622 AGCACAGTGGACCAATATACGATCAACCGCGGTGACCGCGATCGTGGGAGA 1681
```

```
Db 1637 AGCACAGTGGAGCCAAATATACGATCAACCGCGGTGACCGCATCGTGGAGA 1696
QY 1682 TGGGTAAGAGGTACACAGTGTGTTTACGATGCGGATAAATTTGGTTCGGTGTACATTG 1741
Db 1697 CGGTAAGAGGTACACAGTGTGTTTACGATGCGGATAAATTTGGTTCGGTGTACATTG 1756
QY 1742 ATGAGAACCTCTGGAGGGTTTCAGGGCAGACCGTTGTGCGAGACGGAGACGCTTGACA 1801
Db 1757 ATGAGTACTCTTGAAGGGTTTCAGGGCAGACCGTTGTGCGAGACGGAGACGCTTGACA 1816
QY 1802 TCTCCCACTTCTAGCTTGGCGGTATGGAAGGATGATATGCCAAACATAAGCCAGTGA 1861
Db 1817 TTTCCCACTTCTAGCTTGGCAGTTTAAACAGCAGTAAATATGCCAAACATAAGCCAGTGA 1876
QY 1862 CGGTGAATATGTTCTTCTTACCAACCGTCAGCTGAATCCGAGGAGATCAGGACCTTGT 1921
Db 1877 CGGGAATATGTTCTTCTTACCAACCGTCAGCTGAATCCGAGGAGATCAGGACCTTGT 1936
QY 1922 TCTTGAGCCAGGACCTGATTTGGCAGGAAACACATGTCGACGACGACGCGCAGTGT 1981
Db 1937 TCTTGAGCCAGGACCTGATTTGGCAGGAAAGGAATCTGACACAGCAGCAGCGAG 1996
QY 1982 AA 1983
Db 1997 AA 1998

TCU01098 2749 bp mRNA linear INV 29-DEC-1998
Trypanosoma cruzi trans-sialidase homologue mRNA, complete cds.
U01098
U01098.1 GI:624625

SOURCE
ORGANISM
Trypanosoma cruzi
Trypanosoma cruzi
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.
REFERENCE
1 (bases 1 to 2749)
Briones,M.R., Egima,C.M. and Schenkman,S.
Trypanosoma cruzi trans-sialidase gene lacking C-terminal repeats
and expressed in epimastigote forms
Mol. Biochem. Parasitol. 70 (1-2), 9-17 (1995)
95364855
PUBMED 7637718
REFERENCE
2 (bases 507 to 1107)
Briones,M.R., Egima,C.M., Eichinger,D. and Schenkman,S.
Trans-sialidase genes expressed in mammalian forms of Trypanosoma
cruzi evolved from ancestor genes expressed in insect forms of the
parasite
J. Mol. Evol. 41 (2), 120-131 (1995)
95395867
MEDLINE 7666441
PUBMED
REFERENCE
3 (bases 1 to 2749)
Briones,M.R.S.
Direct Submission
Submitted (27-AUG-1993) Marcelo Briones, Escola Paulista de
Medicina, Microbiologia, Imuno e Parasitologia, Rua Botucatu, 862,
8 andar, Sao Paulo, Sao Paulo, Brazil, 04023-062
On Jan 16, 1995 this sequence version replaced gi:430944.
Location/Qualifiers
1. 2749
/organism="Trypanosoma cruzi"
/mol_type="mRNA"
/strain="Y"
/db_xref="taxon:5693"
/dev_stage="epimastigote"
136. 2346
/codon_start=1
/product="trans-sialidase homologue"
/protein_id="AAC98544.1"
/db_xref="GI:624625"

FEATURES
source
CDS
```


/translation="MPRLLLVVLVLPFLNFCNPNVPAHALAPBESSRIELFKRNKSTVP
FEDKGVTERVHSFRPALVNDVGVMAIADARYSNDNSLIDTVVKYSVDDGETW
ETQIGKSVSVSVRVDPVTIVKGNKLVVLVGSYNKSGSYMTWPDGSDWDLIAV
GVVTKTADCKTANI TWGSPVSLKEFPKAYMEGIPTKQFLVGAGVATVNASGNLVVP
VOYTNKKELFSKIPTYSADRGKTNFSEGRSDRGSEPPVVLWEWENKFLINRVDYRR
LVYNSGDKMKPWVEAVVTLSRVWGPSPKSDPGSQSFYATVTEGRVWMLFTHPLNFK
GKWRDLNWLNDQNKQIVNGQVSIGDENSAYSVDLYKDDKLYCLHEINNEVYSLV
FARLVGELMTIKVLSWKNWDSHLSICTPADPATLSSBRGCGPATVTVGLAGFLSD
NATONWEDAYRCVNASTANAEPVNGFKFAGVGAGALWFWQSQGNQNPYHFANYFT
LVASVTIHEVPRAATPLLGASLDSGGKLLGLSYDBKHQWPIYGSPTATPTGSRVEV
NKYHVVLWANKMGSVYIDGVLLKSGSQTVVDEGTPDISHFVIGSYNSNMPTESH
LTKAVNLYNRQLNAKEIRPLFLSQDLGVEEDKSDSDGEDKEQORRERQRTVTV
IRKALLAAMHRQCLLVBSICMPFLWCCTCLSLFLILDMAEEMHYISSYIFMSG"
2715. .2720
polyA signal
polyA_site
2749
/note="17 A nucleotides"

ORIGIN

Query Match		80.7%;	Score 1622.2;	DB 3;	Length 2749;
Best Local Similarity		90.2%;	Pred. No. 0;		
Matches 1735;		Conservative 0;	Mismatches 188;	Indels 0;	Gaps 0;
Qy	62	TGSCACCCGATCGAGCCGAGTTGAGCTGTTTAAGCGGCAAGCTCGAAGTGCCATTG	121		
Db	212	TGGACCCGAATCAGTCGAATTGAATGTTTAAGCGTAAGAATTCGACGGTGCCGTTG	271		
Qy	122	AAAAGGCGGCAAGTCACCGAGCGGTTGTGCACTCGTTCGCGCTCCCGCCCTTGTTA	181		
Db	272	AAGCAAGGCGCAAGTCACCGAGCGGTTGTGCACTCGTTCGCGCTCCCGCCCTTGTTA	331		
Qy	182	ATGTGAGCGGGTGATGGTGGCCATCGCGAGCGCTCGCTACGAAACATCCAAATGACAAC	241		
Db	332	ATGTGATGGGTGATGGTGGCCATTCGACAGCGCTCGCTACAGTCAATCAATGACAAC	391		
Qy	242	CCCTCATTTGATCGGTGGCAAGTACAGCTGACGATGGAGTGGGAGCTGGAGACCCAAA	301		
Db	392	CCCTCATTTGATCGGTGAGTGAAGTACAGCTGACGATGGAGTGGGAGACTGGAGACTCAA	451		
Qy	302	TTGCGATCAAGAACAGTCTGTCATCGTCTGTTCTCGTGTGGTGGATCCACAGTGAATTG	361		
Db	452	TTGGCATCAAGAACAGTCTGTCATCGTCTGTTCTCGTGTGGTGGATCCACAGTGAATTG	511		
Qy	362	TGAAGGCAACAAGCTTTAGCTCTGTTGGAAGTACAAAGTTCGAGAGCTACTGGA	571		
Db	512	TGAAGGCAACAAGCTTTAGCTCTGTTGGAAGTACAAAGTTCGAGAGCTACTGGA	571		
Qy	422	CGTTCGATGCTGATCGAGAGATCGGATATCTGTTGCTGCTGCTGCTGCTGCTGCTGCTG	481		
Db	572	CTTGGCAGCTGATGGAAGCGACTGGGATATCTGATTTGCTGCTGCTGCTGCTGCTGCTG	631		
Qy	482	CCACTGGCGGCGCAAGATAAATCGAGATATCAAAATGGGGGAGCCCGCTGCTACTGAAGG	541		
Db	632	CCACTGGAGAGCGCAAGAACTTGGGAATATCAATGGGGGAGCCCGCTGCTACTGAAGG	691		
Qy	542	AAATTTTCCCGCGGAAATGGAAGGATGACACAAATCAATTTTTCGCGGTGCGAGGTG	601		
Db	692	AAATTTTCCCGGCATACATGGAAGGATACCCCAAGAACAAATTTCTCGCGGTGCGAGGTG	751		
Qy	602	TTGCGATTGGGTGCTCAACCGGAAATCTTGTGTACCTGTGCGAGTTTACGAAACAAAGA	661		
Db	752	TTGCGATTGGGTGCTCAACCGGAAATCTTGTGTACCTGTGCGAGTTTACGAAACATGAAA	811		
Qy	662	AGCAAGTTTTCACAGATCTTCTACTCGGAAGACGAGGCAAGACGTGGAAGTTTGGGG	721		
Db	812	AAGAGCTTTTCTCAAGATCTTTTATTCGGCAGATGAGGCAAGACGTGGAATTTCTCAG	871		
Qy	722	AGGTAGAGGTGATTTTGGTGTCTGAACTGTGGCCCTTGAAGTGGGAGGGAAGCTCA	781		
Db	872	AGGTAGAGGATTTTGGTGTCTCGAACTGTGGTCTTGTAGTGGGAAAAATAAATTC	931		
Qy	782	TCATAAACACTCGAGTTGACTATCGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	841		
Db	932	TCATAAACACTCGAGTTGACTATCGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	991		

Qy	842	ATTCTGTGGGTGAGGGCTGTTCGGCAACGCTCTCAAGTGTGTGGGGCCCTCACCACAAATCGA	901		
Db	992	AACCGTGGGTGAGGGCTGTTCGTCAACGCTCTCGGTGTGTGGGGCCCTCACCACAAATCGG	1051		
Qy	902	ACCAGCCGGCAGCTCAGAGCAGCTTCACTGCGGTGACCATCGAGGGAATGCGTGTATGC	961		
Db	1052	ACCAGCCGGCAGCTCAGAGCAGCTTCACTGCGGTGACCATCGAGGGAATGCGTGTATGC	1111		
Qy	962	TCCTTCACACACCCCGCTGAATTTTAAAGGAAAGTGGCTGCGCGACCGACTGAACCTCTGGC	1021		
Db	1112	TCCTTCACACACCCCGCTGAATTTTAAAGGAAAGTGGCTGCGCGACCGACTGAACCTCTGGC	1171		
Qy	1022	TGACGAGATAACGAGCGCATTTATAACGTTGGGCAAGTATCCATTTGGTGTATGAAAAATTCG	1081		
Db	1172	TGACGAGATAACGAGCGCATTTATAACGTTGGGCAAGTATCCATTTGGTGTATGAAAAATTCG	1231		
Qy	1082	CCTTACAGCTCCGCTCTGTACAAGGATGAAGCTGTACTGTTTGCATGAGATCAACAGTA	1141		
Db	1232	CCTTACAGCTCCGCTCTGTACAAGGATGAAGCTGTACTGTTTGCATGAGATCAACACTA	1291		
Qy	1142	ACGAGGTGTACAGCCTTGTGTTTTCGCGCGCTGTTGGCGAGCTACGAGTCAATTAATCAG	1201		
Db	1292	ACGAGGTGTACAGCCTTGTGTTTTCGCGCGCTGTTGGCGAGCTAATGATTAATTAATCAG	1351		
Qy	1202	TGCTGCAAGTCTCGAAGAAATTTGGGACAGCCACTGTCTCAGCAATTTGCACCCTGCTGATC	1261		
Db	1352	TGCTGCAAGTCTCGAAGAAATTTGGGACAGCCACTGTCTCAGCAATTTGCACCCTGCTGATC	1411		
Qy	1262	CAGCGCTTCTGCTCAGAGCGTGTGTGCTCCGCTGTCACACCGTTGGTCTCTGCTG	1321		
Db	1412	CCGCGCTTCTGCTCAGAGCGTGTGTGCTCCGCTGTCACACCGTTGGTCTCTGCTG	1471		
Qy	1322	GCTTTTGTTCGACAGTGCACCAAAACCGAATGGGAGGATGCGTACCGCTGCGTCAACG	1381		
Db	1472	GCTTTTGTTCGACAGTGCACCAAAACCGAATGGGAGGATGCGTACCGCTGCGTCAACG	1531		
Qy	1382	CAAGCAGCGCAAAATCGGAGAGGTTCCGAAACGGTTTGAAGTTTTCGGGGGTTGGCGGAG	1441		
Db	1532	CAAGCAGCGCAAAATCGGAGAGGTTCCGAAACGGTTTGAAGTTTTCGGGGGTTGGCGGAG	1591		
Qy	1442	GGGCGCTTTCGGCGGTGAGCCAGCGGCGAGATCAACGGTATCACTTTGCAAAACCAAG	1501		
Db	1592	GAGCGCTTTCGGCGGTGAGCCAGCGGCGAGATCAACGGTATCACTTTGCAAAACCAAG	1651		
Qy	1502	CGTTTACGCTGTGTGGGTTCGCTGACGATTCACGAGGTTCCGAGCGTCCGAGTCTCTTGC	1561		
Db	1652	AGTTTACGCTGTGTGGGTTCGCTGACGATTCACGAGGTTCCGAGCGTCCGAGTCTCTTGC	1711		
Qy	1562	TGGGTGCGAGCTTGGATCTTCTTGTGGGCAAAAACTCTTGGGGCTCTCGTACGACGAGA	1621		
Db	1712	TGGGTGCGAGCTTGGATCTTCTTGTGGGCGAAAACTCTTGGGGCTCTCGTACGACGAGA	1771		
Qy	1622	AGCACGAGTGGCAGCCCAATATACGATCAACCGCGGTGACCGCGACCGGATCGTGGGAGA	1681		
Db	1772	AACACGAGTGGCAGCCGATATACGATCAACCGCGGTGACCGCGACCGGATCGTGGGAGG	1831		
Qy	1682	TGGGTGAAGAGTACACGCTGTTCTTACGATGCGAATAAAATTTGGTTTCGGTGTACATTG	1741		
Db	1832	TGAACAGAGATACACGCTGTTCTTACGATGCGAATAAAATTTGGTTCGGTGTACATTG	1891		
Qy	1742	ATGAGAGAACCTCTGAGGGTTTCAAGGCGAGACCGTTGTGCGAGACGGGAGACGCTGACA	1801		
Db	1892	ATGAGAGAACCTCTGAGGGTTTCAAGGCGAGACCGTTGTGCGAGACGGGAGACGCTGACA	1951		
Qy	1802	TCCTCCACTTCTAGCTTGGCGGTATGGAAGGATGATATGCCAACCATTAAGCCACCTGA	1861		
Db	1952	TTTCCCACTTCTAGCTTGGCGGTATGGAAGGATGATATGCCAACCATTAAGCCACCTGA	2011		
Qy	1862	CGGTGAATATGTTCTTTTACAAACCGTTCAGTGAATGCGGAGGATCAGGACCTTGT	1921		
Db	2012	CGCGAAAAACGTTTCTTTTACAAACCGTTCAGTGAATGCGGAGGATCAGGACCTTGT	2071		

QY	1922	TCTTGAGCCAGGACCTGATTGGCACGGAAGCACACATGGGGCAGCAGCAGCGGCAGCAGTG	1981
Db	2072	TCTTGAGCCAGGATCTTGATGGCGTGGAGAGAGACAGGACAGCAGCAGCGGCAGGAGG	2131
QY	1982	AAA	1984
Db	2132	ACA	2134

Search completed: August 8, 2004, 22:34:32
Job time : 12414 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 8, 2004, 17:17:28 ; Search time 1143 Seconds
(without alignments)
7470.585 Million cell updates/sec

Title: US-10-086-913-1

Perfect score: 2010

Sequence: 1 atgggcagcagccatcatca.....cgcccgatccggctgtctaa 2010

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2010	100.0	2010	6	ABA98876 Plasmid e
2	2010	100.0	2010	9	Add06429 DNA encod
3	1924.6	95.8	3183	2	Aax26611 Trypanoso
4	1875.4	93.3	1929	2	Aax26612 Alpha(2-3
5	1753.8	87.3	2133	9	Adc21500 T. cruzi
6	1708.2	85.0	5403	9	Adc21532 T. cruzi
7	498.4	24.8	500	2	Aaq49594 Encodes p
8	498.4	24.8	500	2	Aaq49597 Clone 154
9	484	24.1	500	2	Aaq49595 Clone 121
10	458.4	22.8	500	2	Aaq49596 TCNA tran
11	186.6	9.3	3033	3	Aaz29720 Modified
12	186.6	9.3	3253	3	Aaz29719 T. cruzi
13	179	8.9	2446	2	Aat69167 Trypanoso
14	179	8.9	2446	2	Aax81756 DNA encod
15	67.8	3.4	1068	7	Abv74081 DNA encod
16	67.6	3.4	585	5	Aad16128 Hexa-His-
17	66.2	3.3	467	3	Aaz38274 Human tyr
18	66.2	3.3	654	3	Aaz50043 DNA encod
19	66.2	3.3	1455	5	AAS00250 LFn-Bcl-X
20	66	3.3	1989	3	Aaa49432 Neisseria
21	65.2	3.2	619	3	Aaz50042 DNA encod
22	65	3.2	292	9	Adc19976 pET-28a-c
23	65	3.2	366	5	Aaf25008 Nucleotid

24	65	3.2	543	6	AAS21126	CDNA enco
25	65	3.2	543	6	AAS21122	CDNA enco
26	65	3.2	543	6	AAS21124	CDNA enco
27	65	3.2	543	6	AAS21116	CDNA enco
28	65	3.2	543	6	AAS21118	CDNA enco
29	65	3.2	543	6	AAS21120	CDNA enco
30	65	3.2	552	2	AAV13947	Salmonell
31	65	3.2	681	6	AAS21125	CDNA enco
32	65	3.2	681	6	AAS21121	CDNA enco
33	65	3.2	681	6	AAS21115	CDNA enco
34	65	3.2	682	6	AAS21123	CDNA enco
35	65	3.2	682	6	AAS21117	CDNA enco
36	65	3.2	776	6	AAS21119	CDNA enco
37	65	3.2	900	6	ABV73014	Recombina
38	65	3.2	1067	9	ADB37646	Streptomy
39	65	3.2	1074	9	ADB37648	Streptomy
40	65	3.2	1230	5	AAF25127	Nucleotid
41	65	3.2	1247	2	AAX29725	Insert fr
42	65	3.2	1320	8	AAL56894	Partial S
43	65	3.2	1435	7	ACC44570	pNOV4835
44	65	3.2	1436	7	ACC44571	pNOV4836
45	65	3.2	1707	3	AAA49438	Candida g

ALIGNMENTS

RESULT 1

ABA98876
ID ABA98876 standard; DNA; 2010 BP.

XX ABA98876;

DT 01-JUL-2002 (first entry)

DE Plasmid encoding the catalytic trans-sialidase unit of T. cruzi.

XX Mycoplasma associated disease; cell proliferation; trans-sialidase;
KW enzyme; atherosclerotic vascular disease; malignancy; sialic acid;
KW antiatherosclerotic; antibacterial; antiviral; anti-HIV; cytostatic;
KW vasotrophic; ovarian carcinoma; breast cancer; prostate cancer;
KW colon cancer; lung cancer; leukaemia; HIV; human immunodeficiency virus;
KW chlamydia; PCR primer; ds.
XX Trypanosoma cruzi.
OS Synthetic.

XX Key Location/Qualifiers
FT CDS 1..2010

FT /product= "catalytic trans-sialidase unit"

W0200202050-A2.

XX 10-JAN-2002.

XX 03-JUL-2001; 2001WO-BR000083.

PR 03-JUL-2000; 2000BR-00002989.

XX (HIGU)/ HIGUCHI M D L.

PA (SCHE/) SCHENKMAN S.

PI Higuchi MDL, Schenkman S;

QY 721 GAGGTAGGAGTGAATTTGGCTGCTCTGAACCTGTGCGCTTGGAGTGGAGGGAGCTC 780
DB 721 GAGGTAGGAGTGAATTTGGCTGCTCTGAACCTGTGCGCTTGGAGTGGAGGGAGCTC 780
QY 781 ATCATAAACACTCGAGTTGATATCCGCCCGCTCTGGTGTACAGTCCAGTGCATGGGG 840
DB 781 ATCATAAACACTCGAGTTGATATCCGCCCGCTCTGGTGTACAGTCCAGTGCATGGGG 840
QY 841 AATTCTGGTGGAGCTGTGGCAGCTCTCACTGCTGTGGGGCCCTTCACAAAATCG 900
DB 841 AATTCTGGTGGAGCTGTGGCAGCTCTCACTGCTGTGGGGCCCTTCACAAAATCG 900
QY 901 AACAGCCCGGAGTCAGAGAGCTTCACTGCGCTGACCATCGAGGGAATCGTGTATG 960
DB 901 AACAGCCCGGAGTCAGAGAGCTTCACTGCGCTGACCATCGAGGGAATCGTGTATG 960
QY 961 CTCCTTACACACCCCGCTGAATTTTAAAGGAAGGTGGCTGCGCAGCCGACTGAACCTCTGG 1020
DB 961 CTCCTTACACACCCCGCTGAATTTTAAAGGAAGGTGGCTGCGCAGCCGACTGAACCTCTGG 1020
QY 1021 CTGACGGATAACAGCGCATTTTAAAGCTGTGGCAAGTATCCATTGGTGATGAAATTC 1080
DB 1021 CTGACGGATAACAGCGCATTTTAAAGCTGTGGCAAGTATCCATTGGTGATGAAATTC 1080
QY 1081 GCCTACAGCTCCGTCTGTACAGGATGATAGCTGTACTGTTTGCATGAGATCAACAGT 1140
DB 1081 GCCTACAGCTCCGTCTGTACAGGATGATAGCTGTACTGTTTGCATGAGATCAACAGT 1140
QY 1141 AACGAGGTGTACAGCTTGTGTTTTCGGCCCTGTGGTGGAGCTACGGATCATTAATCA 1200
DB 1141 AACGAGGTGTACAGCTTGTGTTTTCGGCCCTGTGGTGGAGCTACGGATCATTAATCA 1200
QY 1201 GTGCTGCAGTCTCTGAAGAAATTTGGACAGCAACCTGTCCAGCATTTGCACCCCTGCTGAT 1260
DB 1201 GTGCTGCAGTCTCTGAAGAAATTTGGACAGCAACCTGTCCAGCATTTGCACCCCTGCTGAT 1260
QY 1261 CCAGCCGCTTGTGTCTAGAGGTGTTGTGGTCCCGCTGTACCAAGTGTGTTGTT 1320
DB 1261 CCAGCCGCTTGTGTCTAGAGGTGTTGTGGTCCCGCTGTACCAAGTGTGTTGTT 1320
QY 1321 GCTTTTGTGCGCAGTGCACCAACCAAGTGGGAGGATGGTACCGCTGCTCAAC 1380
DB 1321 GCTTTTGTGCGCAGTGCACCAACCAAGTGGGAGGATGGTACCGCTGCTCAAC 1380
QY 1381 GCAAGCAGGCAATGCGGAGGGTTCCGAACGGTTTGAAGTTTGGGGGGTTGGCGA 1440
DB 1381 GCAAGCAGGCAATGCGGAGGGTTCCGAACGGTTTGAAGTTTGGGGGGTTGGCGA 1440
QY 1441 GGGGGCTTTGGCCGGTGAGCCAGCAGGGGCAAGATCAACGGTATCACTTTGCAACAC 1500
DB 1441 GGGGGCTTTGGCCGGTGAGCCAGCAGGGGCAAGATCAACGGTATCACTTTGCAACAC 1500
QY 1501 GCGTTACGCTGTGGCTGGTGACGATTCACGAGTTCCGAGCGTCCGAGTCCCTTTG 1560
DB 1501 GCGTTACGCTGTGGCTGGTGACGATTCACGAGTTCCGAGCGTCCGAGTCCCTTTG 1560
QY 1561 CTGGTTCGAGCTGGACTCTTCTGTGGGCAAAAACCTCTGGGGCTCTCGTACGACGAG 1620
DB 1561 CTGGTTCGAGCTGGACTCTTCTGTGGGCAAAAACCTCTGGGGCTCTCGTACGACGAG 1620
QY 1621 AAGCAGGTGGCAGCAATATACGATCAACCGGTGACCGCGGATCGTGGGAG 1680
DB 1621 AAGCAGGTGGCAGCAATATACGATCAACCGGTGACCGCGGATCGTGGGAG 1680
QY 1681 ATGGTTAAGAGGTACCACTGTTCTTACGATGGCAATAAATTTGGTTCCGTTGACATT 1740
DB 1681 ATGGTTAAGAGGTACCACTGTTCTTACGATGGCAATAAATTTGGTTCCGTTGACATT 1740
QY 1741 GATGGAGAACTCTGGAGGGTTGAGGGAGACCGTTGTGCCAGACGGGAGGACGCTGAC 1800
DB 1741 GATGGAGAACTCTGGAGGGTTGAGGGAGACCGTTGTGTGCCAGACGGGAGGACGCTGAC 1800

QY 1801 ATCTCCCACTTCTACGTTGGCGGATGGAAGAGTGATATGCCAACCAATAAGCCACGTG 1860
DB 1801 ATCTCCCACTTCTACGTTGGCGGATGGAAGAGTGATATGCCAACCAATAAGCCACGTG 1860
QY 1861 ACGGTGAATAATGTTCTTCTTACACCGTCACTGAATGCCAGGAGATCAGGACCTTG 1920
DB 1861 ACGGTGAATAATGTTCTTCTTACACCGTCACTGAATGCCAGGAGATCAGGACCTTG 1920
QY 1921 TTCTTGAGCCAGGACCTGATTTGGCAGGAAGCACACATGGGACGAGCAGCGCAGCAGT 1980
DB 1921 TTCTTGAGCCAGGACCTGATTTGGCAGGAAGCACACATGGGACGAGCAGCGCAGCAGT 1980
QY 1981 GAAAGAGTACGCCCGGATCCGCTCTCTAA 2010
DB 1981 GAAAGAGTACGCCCGGATCCGCTCTCTAA 2010

RESULT 3
AAX26611

ID AAX26611 standard; DNA; 3183 BP.

XX AAX26611;

XX AC AAX26611;

XX XX 15-JUN-1999 (first entry)

XX XX Trypanosoma cruzi

XX DE Alpha (2-3) trans-sialidase; sialyl-oligosaccharide; sialyllactose;
XX KW cheese processing waste strain; (2-3) sialyllactose; gastric ulcer;
XX KW duodenal ulcer; arthritis; enterotoxin; ss.
XX OS Trypanosoma cruzi.

XX XX WO9908511-A1.

XX XX 25-FEB-1999.

XX PF 13-AUG-1998; 98WO-US016756.

XX XX 14-AUG-1997; 97US-00911393.

XX XX (NEOS-) NEOSE TECHNOLOGIES INC.

XX XX Pelletier M, Barker WA, Hakes DJ, Zopf DA;

XX XX WPI: 1999-190079/16.

XX XX P-PSDB; AAY01540.

XX XX Production of sialyl-oligosaccharides, particularly sialyl-lactose - by
XX XX treating a dairy source such as a cheese processing waste stream with an
XX XX alpha (2-3) trans-sialidase.

XX XX Disclosure; Fig 1; 84pp; English.

XX XX The present sequence encodes Trypanosoma cruzi alpha (2-3) trans-
XX XX sialidase. The protein is used in the method of the invention to produce
XX XX sialyl-oligosaccharides, particularly sialyllactose, which are produced
XX XX by treating a dairy source such as a cheese processing waste strain with
XX XX an alpha (2-3) trans-sialidase. The method can be used for producing
XX XX sialyl-oligosaccharides, such as (2-3) sialyllactose for pharmaceutical
XX XX use. (2-3) sialyllactose has been shown to neutralise enterotoxins of
XX XX various pathogenic microbes including E. coli, Vibrio cholerae and
XX XX Salmonella. It has also been shown that alpha (2-3) (2-3) sialyllactose
XX XX (alpha-Neu5Ac-(2-4)-Gal-beta-(1-4)-Glc) interferes with colonisation of
XX XX Helicobacter pylori and thereby prevents or inhibits gastric and duodenal
XX XX ulcers. (2-3) sialyllactose has additionally been proposed to inhibit
XX XX immune complex formation by disrupting occupancy of the Fc carbohydrate
XX XX binding site on IgG and to be useful in treating arthritis

XX XX Sequence 3183 BP; 673 A; 964 C; 897 G; 649 T; 0 U; 0 Other;

XX XX Query Match 95.8%; Score 1924.6; DB 2; Length 3183;

XX XX Best Local Similarity 99.5%; Pred. No. 0;

	Matches	1930;	Conservative	0;	Mismatches	9;	Indels	0;	Gaps	0;
Qy	62	TGGCA	CCCGGATCGAGCGGAGTTCGAGCTGTTTAAAGCGGCAAGCTCGAAGGTGCAATTG	121						
Db	98	TGGCAC	CCGGATCGAGCGGAGTTCGAGCTGTTTAAAGCGGCAAGCTCGAAGGTGCAATTG	157						
Qy	122	AAAAAGGGCGGCAAAAGTCAACGAGCGGGTGTTCACACTGTTTCGCGCTCCCGCCCTTGTGA	181							
Db	158	AAAAAGGGCGGCAAAAGTCAACGAGCGGGTGTTCACACTGTTTCGCGCTCCCGCCCTTGTGA	217							
Qy	182	ATGTGGACCGGGTGATGTGTTCCCATCGCGACGCTCGCTACGAAACATCCAATGACAACT	241							
Db	218	ATGTGGACCGGGTGATGTGTTCCCATCGCGACGCTCGCTACGAAACATCCAATGACAACT	277							
Qy	242	CCCTCATTTGATACGGTGGCGAAGTACAGCGTGGACGATGGGGAGACGTGGGAGACCCAAA	301							
Db	278	CCCTCATTTGATACGGTGGCGAAGTACAGCGTGGACGATGGGGAGACGTGGGAGACCCAAA	337							
Qy	302	TTGCCATCAAGAA CAGTGTGTGCATCGTCTGTTTCTCGTGTGTGTGATCCACAGTGATG	361							
Db	338	TTGCCATCAAGAA CAGTGTGTGCATCGTCTGTTTCTCGTGTGTGTGATCCACAGTGATG	397							
Qy	362	TGAAGGCGCAACAAGCTTTACGTCTCGTTGGAAGCTACACAGTTTCGAGGAGCTACTGGA	421							
Db	398	TGAAGGCGCAACAAGCTTTACGTCTCGTTGGAAGCTACACAGTTTCGAGGAGCTACTGGA	457							
Qy	422	CGTCGCATGTGTGATCGGAGACACTGGGATATTTCTGCTTGGCGTTGGTGAGGTCAACGA	481							
Db	458	CGTCGCATGTGTGATCGGAGACACTGGGATATTTCTGCTTGGCGTTGGTGAGGTCAACGA	517							
Qy	482	CCACTCGGGCGGCAAGATAACTCGGAGTATCAAATGGGGAGACCCCGTGTCACTGAAGG	541							
Db	518	CCACTCGGGCGGCAAGATAACTCGGAGTATCAAATGGGGAGACCCCGTGTCACTGAAGG	577							
Qy	542	AAATTTTCCCGCGGGAATGGAAGGAATGCACACAAATCAATTTCTTTGGCGGTGCAGGTG	601							
Db	578	AAATTTTCCCGCGGGAATGGAAGGAATGCACACAAATCAATTTCTTTGGCGGTGCAGGTG	637							
Qy	602	TTGCCATTGTGGCGTCCAA CGGGAACTCTTGTGTACCTGTGCAGGTGTACGAA CAAAAGA	661							
Db	638	TTGCCATTGTGGCGTCCAA CGGGAACTCTTGTGTACCTGTGCAGGTGTACGAA CAAAAGA	697							
Qy	662	AGCAAGTTTTTTTCCAAGATCTTCTACTCGGAAGACGAGGGCAAGCTGTGAAGTTTGGGG	721							
Db	698	AGCAAGTTTTTTTCCAAGATCTTCTACTCGGAAGACGAGGGCAAGCTGTGAAGTTTGGGG	757							
Qy	722	AGGTAGAGTGATTTTGGTGTCTTGAA CCTGTGGCCCTTGAGTGGGAGGGGAGGTCA	781							
Db	758	AGGTAGAGTGATTTTGGTGTCTTGAA CCTGTGGCCCTTGAGTGGGAGGGGAGGTCA	817							
Qy	782	TCATAAACACTCGAGTTGACTATTCGCGCGCTCTGTGTACGAGTCCAGTGCATGAGGGGA	841							
Db	818	TCATAAACACTCGAGTTGACTATTCGCGCGCTCTGTGTACGAGTCCAGTGCATGAGGGGA	877							
Qy	842	ATTCTGTGGGTGGAGGCTGTTCGGCACGCTCTCACGTGTGTGGGGCCCTCACCAAAAATCGA	901							
Db	878	ATTCTGTGGGTGGAGGCTGTTCGGCACGCTCTCACGTGTGTGGGGCCCTCACCAAAAATCGA	937							
Qy	902	ACGAGCCCGGAGTTCAGAGCAGCTTCTACCTGCGTGA CCACTCGAGGGGAATCGCTGTTATGC	961							
Db	938	ACGAGCCCGGAGTTCAGAGCAGCTTCTACCTGCGTGA CCACTCGAGGGGAATCGCTGTTATGC	997							
Qy	962	TCCTTCACACACCCGCTCAATTTTAAAGGGAGGTGGCTGCGGACCGACTCAACCTCTGGC	1021							
Db	998	TCCTTCACACACCCGCTCAATTTTAAAGGGAGGTGGCTGCGGACCGACTCAACCTCTGGC	1057							
Qy	1022	TGACGGATACACAGCGCATTTATAACCGTTGGGCAAGTATCCATTTGGTGTATGAAAATTCGG	1081							
Db	1058	TGACGGATACACAGCGCATTTATAACCGTTGGGCAAGTATCCATTTGGTGTATGAAAATTCGG	1117							
Qy	1082	CCTACAGCTCCGTCCTGTACAAGGATGATTAAGCTGTACTGTTTGCATGTAGATCAACAGTA	1141							
Db	1118	CCTACAGCTCCGTCCTGTACAAGGATGATTAAGCTGTACTGTTTGCATGTAGATCAACAGTA	1177							

QY	1142	ACGAGGTGTACAGCCTTGTGTCGGCCCTGTTGGCGAGACTACGATCATTTAAATCAG	1201
DB	1178	ACGAGGTGTACAGCCTTGTGTCGGCCCTGTTGGCGAGACTACGATCATTTAAATCAG	1237
QY	1202	TGCTGCAGTCTCTGGAAGAAATTGGGACAGCACCTGTCCAGCATTTTGACCCCTGCTGATC	1261
DB	1238	TGCTGCAGTCTCTGGAAGAAATTGGGACAGCACCTGTCCAGCATTTTGACCCCTGCTGATC	1297
QY	1262	CAGCCGCTTCTGTCTAGAGCGTGTGTCGTCGCTGTTCACACGGTTGGTCTTGTGTTG	1321
DB	1298	CAGCCGCTTCTGTCTAGAGCGTGTGTCGTCGCTGTTCACACGGTTGGTCTTGTGTTG	1357
QY	1322	GCTTTTGTGTCACAGTGCCACCAAAACCGAATGGGAGATGCGTACCGCTCGTCAACG	1381
DB	1358	GCTTTTGTGTCACAGTGCCACCAAAACCGAATGGGAGATGCGTACCGCTCGTCAACG	1417
QY	1382	CAAGCAGCGCAAAATCGCGAGAGGTTCCGAACGGTTTGAAGTTTGGCGGGTTTGGCGAG	1441
DB	1418	CAAGCAGCGCAAAATCGCGAGAGGTTCCGAACGGTTTGAAGTTTGGCGGGTTTGGCGAG	1477
QY	1442	GGGCGCTTTGGCCGGTGAGCCAGAGGGGCGAGAATCAACGGTATCATTTTGCAAAACGAG	1501
DB	1478	GGGCGCTTTGGCCGGTGAGCCAGAGGGGCGAGAATCAACGGTATCATTTTGCAAAACGAG	1537
QY	1502	CGTTACGCTGGTGGCGTCCGTGACAGATTCAACGAGTTTCCGAGCGTCGCGAGTCCCTTGC	1561
DB	1538	CGTTACGCTGGTGGCGTCCGTGACAGATTCAACGAGTTTCCGAGCGTCGCGAGTCCCTTGC	1597
QY	1562	TGGGTGGAGCTGGAATCTTTCTGGTGGCAAAAATCTCTTGGGGCTCTCGTACGACGAGA	1621
DB	1598	TGGGTGGAGCTGGAATCTTTCTGGTGGCAAAAATCTCTTGGGGCTCTCGTACGACGAGA	1657
QY	1622	AGCACAGTGGCAGCCAATATACGGATCAACGCGGTGACGCCGACCGGATCTGGGAGAGA	1681
DB	1658	AGCACAGTGGCAGCCAATATACGGATCAACGCGGTGACGCCGACCGGATCTGGGAGAGA	1717
QY	1682	TGGGTAAAGAGTACCACTGGTCTTACGATGGCGAATAAAATGGTTCCGTTGTCATTTG	1741
DB	1718	TGGGTAAAGAGTACCACTGGTCTTACGATGGCGAATAAAATGGTTCCGTTGTCATTTG	1777
QY	1742	ATGAGAACTCTTGAGGGTTTCAGGCGAGACCGTTGTGCCAGACGGGAGGACCCCTGACA	1801
DB	1778	ATGAGAACTCTTGAGGGTTTCAGGCGAGACCGTTGTGCCAGACGGGAGGACCCCTGACA	1837
QY	1802	TCTCCCACTTCTACGTTGGCGGGTATGGAAAGGAGTGATATGCCAACCATTAAGCCACGTGA	1861
DB	1838	TCTCCCACTTCTACGTTGGCGGGTATGGAAAGGAGTGATATGCCAACCATTAAGCCACGTGA	1897
QY	1862	CGGTGAATAATGTTCTTCTTTACAACCGTCAGCTGAATGCCGAGGAGATCAGGACCTTGT	1921
DB	1898	CGGTGAATAATGTTCTTCTTTACAACCGTCAGCTGAATGCCGAGGAGATCAGGACCTTGT	1957
QY	1922	TCTTGACCCAGGACCTGATTTGGCACGGAAGCACACATGGGCACGACGCGGACGAGTG	1981
DB	1958	TCTTGACCCAGGACCTGATTTGGCACGGAAGCACACATGGGCACGACGCGGACGAGTG	2017
QY	1982	AAAGAAGTACGCCCGGATC	2000
DB	2018	CCACAGTACGCCCTCAAC	2036

RESULT 4

AAX26612

ID AAX26612 standard; DNA; 1929 BP.

XX
 XXXXX

AC AAX26612;

XX

DT 15-JUN-1999 (first entry)

XX

DE Alpha(2-3) trans-sialidase sequence lacking amino acid repeats.

XX

KW	cheese processing waste strain; (2-3)sialyllactose; gastric ulcer;
KW	duodenal ulcer; arthritis; enterotoxin; ss.
XX	
OS	Trypanosoma cruzi.
XX	
PN	WO9908511-A1.
XX	
PD	25-FEB-1999.
XX	
PF	13-AUG-1998; 98WO-US016756.
XX	
PR	14-AUG-1997; 97US-00911393.
XX	
PA	(NEOS-) NEOSE TECHNOLOGIES INC.
PI	Pelletier M, Barker WA, Hakes DJ, Zopf DA;
XX	
DR	WPI; 1999-190079/16.
DR	P-PSDB; AAY01541.
XX	
PT	Production of sialyl-oligosaccharides, particularly sialyl-lactose - by
PT	treating a dairy source such as a cheese processing waste stream with an
PT	alpha (2-3) trans-sialidase.
XX	
PS	Disclosure; Fig 3; 84pp; English.
XX	
CC	The present sequence encodes a functional Trypanosoma cruzi alpha (2-
CC	3)trans-sialidase which lacks amino acid repeats. The protein is used in
CC	the method of the invention to produce sialyl-oligosaccharides,
CC	particularly sialyllactose, which are produced by treating a dairy source
CC	such as a cheese processing waste strain with an alpha (2-3) trans-
CC	sialidase. The method can be used for producing sialyl-oligosaccharides,
CC	such as (2-3)sialyllactose for pharmaceutical use. (2-3)sialyllactose
CC	has been shown to neutralise enterotoxins of various pathogenic microbes
CC	including E. coli, Vibrio cholerae and Salmonella. It has also been shown
CC	that alpha(2-3)(2-3)sialyllactose (alpha-Neu5Ac-(2-3)-Gal-beta-(1-4)-Glc)
CC	interferes with colonisation of Helicobacter pylori and thereby prevents
CC	or inhibits gastric and duodenal ulcers. (2-3)sialyllactose has
CC	additionally been proposed to inhibit immune complex formation by
CC	disrupting occupancy of the Fc carbohydrate binding site on IgG and to be
CC	useful in treating arthritis
XX	
SQ	Sequence 1929 BP; 442 A; 463 C; 590 G; 434 T; 0 U; 0 Other

Seq	Sequence	1929 BP;	442 A;	463 C;	590 G;	434 T;	0 U;	0 Other;	
	Query Match	93.3%;	Score	1875.4;	DB	2;	Length	1929;	
	Best Local Similarity	98.4%;	Pred. No.	0;					
	Matches 1894;	Conservative	0;	Mismatches	31;	Indels	0;	Gaps	0;
QY	62	TGCGACCCGGATCGAGCCGAGTTGAGCTGTGTTTAAGCGGCAAAAGCTCGAAGGTGCCATTGTG	121						
Db	5	TGCGACCCGGATGAGAGCCGAGTTGAGCTGTGTTTAAGCGGCAAAAGCTCGAAGGTGCCATTGTG	64						
QY	122	AAAGGGCGGCAAAAGTCAACGAGCGGGTGTGCCACTCGTTCCGCTCTCCGCGCCTTGTTA	181						
Db	65	AAAGGACGGCAAAAGTCAACGAGCGGGTGTGCCACTCGTTCCGCTCTCCGCGCCTTGTTA	124						
QY	182	ATGTGGACGGGGTGATGTGTTCCCATCGGGAGCGTCTGCTACGAAAATCCAATGACAAC	241						
Db	125	ATGTGGACGGGGTGATGTGTTCCCATCGGGAGCGTCTGCTACGAAAATCCAATGACAAC	184						
QY	242	CCCTCATTTGATACGGTGTGCGAAGTACAGCGTGGACGATGGGAGAGCGTGGGAGACCCAAA	301						
Db	185	CCCTCATTTGATACGGTGTGCGAAGTACAGCGTGGACGATGGGAGAGCGTGGGAGACCCAAA	244						
QY	302	TTGCCATCAAGAACAGTCGTGCATCGTCTCTGTTCTCGTGTGGATCCACAGTGATTG	361						
Db	245	TTGCCATCAAGAACAGTCGTGCATCGTCTCTGTTCTCGTGTGGATCCACAGTGATTG	304						
QY	362	TGAAGGGCAACAAGCTTTACGTCCTCGTTTGGAAAGCTACAAAGTTCGAGGAGCTACTCGA	421						
Db	305	TGAAGGGCAACAAGCTTTACGTCCTCGTTTGGAAAGCTACAAAGTTCGAGGAGCTACTCGA	364						
QY	422	CGTCGCATGTTGATGCGAGAGACTGGGATATCTCGTTCGGTGGTGAAGTCACGAAGT	481						


```
Db 1445 CGTTCACCGTGGTGGCGCTCGGTGACGATTCACGAGGTTCCGAGGTCGCGAGTCCCTTGGC 1504
Qy 1562 TGGGTGCGAGCCTGGACTCTTCTGTTGGCGCAAAAACCTCTCTGGGGCTCTCGTACGACGAGA 1621
Db 1505 TGGGTGCGAGCCTGGACTCTTCTGTTGGCGCAAAAACCTCTCTGGGGCTCTCGTACGACGAGA 1564
Qy 1622 AGACACGTTGGCGAGCCAAATATACGGATCAACGCCGGTGAAGCCGATCGTGGGAGA 1681
Db 1565 GGCACCAGTGGCGAGCCAAATATACGGATCAACGCCGGTGAAGCCGATCGTGGGAGA 1624
Qy 1682 TGGGTAAGAGGTACCAAGCTGTTCTTACGATGCGGAATAAAATGGTTGCGGTGACATTG 1741
Db 1625 TGGGTAAGAGGTACCAAGCTGTTCTTACGATGCGGAATAAAATGGTTGCGGTGACATTG 1684
Qy 1742 ATGGAGAACCTCTGGAGGGTTCAGGGCAGACCCGTTGTGCCAGACGGGAGACGCTGACA 1801
Db 1685 ATGGAGAACCTCTGGAGGGTTCAGGGCAGACCCGTTGTGCCAGACGGGAGACGCTGACA 1744
Qy 1802 TCTCCCACTCTACGTTGGCGGTATGGAAGAGTGATATGCCAACCATTAAGCCAGTGA 1861
Db 1745 TCTCCCACTCTACGTTGGCGGTATGGAAGAGTGATATGCCAACCATTAAGCCAGTGA 1804
Qy 1862 CGGTGAATATGTTCTTTTACACCGTCAGCTGAATGCGAGGAGTACGAGCCTTGT 1921
Db 1805 CGGTGAATATGTTCTTTTACACCGTCAGCTGAATGCGAGGAGTACGAGCCTTGT 1864
Qy 1922 TCTTGAGCCAGGACCTGATTGGCACGGAAGCACACATGGCGAGCAGCGGAGCAGTG 1981
Db 1865 TCTTGAGCCAGGACCTGATTGGCACGGAAGCACACATGGCGAGCAGCGGAGCAGTG 1924
Qy 1982 AAAGA 1986
Db 1925 CTTGA 1929

RESULT 5
ADC21500
ID ADC21500 standard; DNA; 2133 BP.
XX
AC ADC21500;
XX
XT 18-DEC-2003 (first entry)
XX
DE T. cruzi trans-sialidase gene, TS, clone 19Y.
XX
KW Trans-sialidase; gene; TS; ds; neurotrophic peptide; interleukin-6;
KW IL-6 secretion inducing peptide; neuron; glial cell; trophic support;
KW ciliary neurotrophic factor; CNTF; leukaemia inhibitory factor; LIF;
KW amyotrophic lateral sclerosis; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; Chagas' disease; peripheral neuropathy; palsy;
KW multiple sclerosis; stroke; brain trauma; spinal cord trauma;
KW peripheral nerve trauma.
XX
OS Trypanosoma cruzi.
XX
Key Location/Qualifiers
FH 136..2133
FT CDS
FT /tag= b
FT /product= "Trans-sialidase"
FT /partial
FT /note= "No stop codon shown"
FT sig_peptide 136..369
FT /tag= a
FT mat_peptide 370..2133
FT /tag= c
FT /note= "Mature trans-sialidase"
XX
XX US2002137667-A1.
XX
XX 26-SEP-2002.
XX
XX 20-DEC-2000; 2000US-00745008.
XX
XX
```

```
PR 20-DEC-1999; 99US-0172881P.
XX
PA (TUFT ) UNIV TUFTS.
XX
PI Chuenkova M, Pereira MA;
XX
XX WPI; 2003-786654/74.
DR P-PSDB; ADC21501.
XX
PT T. cruzi trans-sialidase derived neurotrophic and interleukin-6 secretion
PT inducing peptides, useful in the treatment of neuronal degeneration
PT caused for example by Chagas' disease.
XX
PS Example 1; SEQ ID NO 1; 79pp; English.
XX
CC The invention relates to a T. cruzi trans-sialidase (TS) derived
CC neurotrophic peptide appearing as ADC21513 or ADC21511 called C44 and
CC C14, or its variant, and an interleukin-6 (IL-6) secretion inducing
CC peptide appearing as ADC21531 called TR-1 (terminal repeat 1). Also
CC included are a composition comprising the peptides, fusion protein
CC comprising the peptides and a fusion partner, a composition comprising
CC the fusion protein and a physiological acceptable carrier, providing
CC trophic support for neurons or glial cells in a mammal (comprising
CC administering a therapeutically effective amount of T. cruzi trans-
CC sialidase (TS) or a neurotrophic variant) and stimulating (M3) the
CC secretion of IL-6 (interleukin-6) in a mammal comprising administering TS
CC or an IL-6 inducing variant. The fusion partner comprises a mammalian
CC neurotrophic factor which is ciliary neurotrophic factor (CNTF) or
CC leukaemia inhibitory factor (LIF). The peptides are useful in providing
CC trophic support for neurons and glial cells in a mammal suffering a
CC condition selected from: amyotrophic lateral sclerosis, Alzheimer's
CC disease, Parkinson's disease, Huntington's disease, Chagas' disease,
CC peripheral neuropathy, palsies, multiple sclerosis, stroke, brain trauma,
CC spinal cord trauma and peripheral nerve trauma, and in stimulating the
CC secretion of IL-6. The present sequence encodes trans sialidase clone
CC 19Y.
XX
SQ Sequence 2133 BP; 485 A; 529 C; 635 G; 484 T; 0 U; 0 Other;
Query Match 87.3%; Score 1753.8; DB 9; Length 2133;
Best Local Similarity 94.3%; Pred. No. 0;
Matches 1831; Conservative 0; Mismatches 107; Indels 3; Gaps 1;
Qy 22 CATCATCACAGCAGCGCGCTGGTGGCGCGGAGCCATATGACCCGATCGAGCCGA 81
Db 193 CTTCTTCTTGGCATCTGCCCGAGCGCGCTGACGCTTGGCACCGGATCGAGCCGA 252
Qy 82 GTTGAGCTGTTTAAGCGGCAAGCTCGAAGTGCCTTTGAA---AAGGGCGCAAGTC 138
Db 253 GTTGAGCTGTTTAAGCGTAAGAATTGAGCGGTCCCGTTTGAAGACAAGGCGGCAAGTC 312
Qy 139 ACCGAGCGGGTGTCCACTCGTTCCGCTCCCGCTCCCGCTTGTAAATGTGACGGGTGATG 198
Db 313 ACCGAGCGGGTGTCCACTCGTTCCGCTCCCGCTCCCGCTTGTAAATGTGACGGGTGATG 372
Qy 199 GTTGCCATCGCGACGCTCGCTACGAACAATCAATGACAACCTCCCTCATTTGATACGGTG 258
Db 373 GTTGCCATCGCGACGCTCGCTACGACACATCAATGACAACCTCCCTCATTTGATACGGTG 432
Qy 259 GCGAAGTACAGCGTGAACGATGGGAGACGTGGAGACCCCAATTCGATCAAGNACAGT 318
Db 433 GCGAAGTACAGCGTGAACGATGGGAGACGTGGAGACCCCAATTCGATCAAGNACAGC 492
Qy 319 CGTGCACTCGTCTGTTTCTCGTGTGGTGGATCCCACTGATGTTGTAAGGCGCAACAGCTT 378
Db 493 CGTGATCGTCTGTTTCTCGTGTGGTGGATCCCACTGATGTTGTAAGGCGCAACAGCTT 552
Qy 379 TACGTCTCTGGTGGAAAGCTACAAAGTTCGAGAGCTACTGGAAGTGTGATGTCGCG 438
Db 553 TACGTCTCTGGTGGAAAGCTACTATAGTTCGAGAAGCTACTGTCGTCGTCATGTCG 612
Qy 439 AGAGACTGGGATATCTGCTTGGCTTGGTGGAGTCAAGAGTCCACTCGGCGGCGCAAG 498
|||||
```

Db 613 AGAGACTGGGATATTCCTGCTTCCGCTTGGTGAGGTCACGAACTCCACTGCGGGCGCAAG 672
QY 499 ATAACTGCGAGTATCAAAATGGGGAGCCCGGTGTCACCTGAAGGAATTTTCCCGCGGAA 558
Db 673 ATAACTGCGAGTATCAAAATGGGGAGCCCGGTGTCACCTGAAGGAATTTTCCCGCGAA 732
QY 559 ATGGAAGGAATGACACAAATCAATTTCTTGGCGGTGCAGGTGTTGCCATTTGTCGCTCC 618
Db 733 ATGGAAGGCATGACACAAATCAATTTCTTGGCGCGGGTGTGGCATTTGTAGCGTCC 792
QY 619 AACGGGAATCTTGTGTACCTGTGACCTGTGACGTTACGAACAAAGAAAGCAAGATTTTCCAA 678
Db 793 AACGGGAATCTTGTGTACCTGTGACCTGTGACGTTACGAACAAAGAAAGCAAGATTTTCCAA 852
QY 679 ATCTTCTACTCCGAAGACGAGGCGCAAGACGTGGAAGTTTGGGAGGGTAGAGTGAATTT 738
Db 853 ATCTTCTACTCCGAAGATGATGGAAGACGTGGAAGTTTGGGAGGGTAGAGTGAATTT 912
QY 739 GGCTGCTTGAACTGTGGCCCTTGAGTGGGAGGGAAGCTCATATAAAACACTCGAGTT 798
Db 913 GGCTGCTTGAACTGTGGCCCTTGAGTGGGAGGGAAGCTCATATAAAACACTCGAGTT 972
QY 799 GACTATCGCCCGCTGCTGCTGACGNTCCAGTGCATGCGGAATTTGTTGGGTGAGGCT 858
Db 973 GACTGGAACGCGGCTGCTGCTGACGNTCCAGTGCATGCGGAATTTGTTGGGTGAGGCT 1032
QY 859 GTCGACGCTCTCACGCTGTGTGGGGCCCTCACCAAAATCGAACACGAGCCGCGAGTCAG 918
Db 1033 GTCGGAACGCTCTGCGGTGTGTGGGGCCCTCACCAAAATCGAACACGAGCCGCGAGTCAG 1092
QY 919 AGCAGCTTCACTGCGGTGACCATCGAGGGAATGCGTGTATGCTCTTCAACACCCGCTG 978
Db 1093 AGCAGCTTCACTGCGGTGACCATCGAAGGAATGCGTGTATGCTCTTCAACACCCGCTG 1152
QY 979 AATTTTAAGGAAGTGGCTGCGGACCGACTGAACTCTGGCTGACGGATACACCGC 1038
Db 1153 AATTTTAAGGAAGTGGCTGCGGACCGACTGAACTCTGGCTGACGGATACACCGC 1212
QY 1039 ATTTATTAACGTTGGGCAAGTATCCATTTGGTGAATGAAATTCGCGCTACAGCTCCGCTCG 1098
Db 1213 ATTTATTAACGTTGGGCAAGTATCCATTTGGTGAATGAAATTCGCGCTACAGCTCCGCTCG 1272
QY 1099 TACAAGATGATAGCTGTAAGTGTTCATGAGATCAACAGTAAACGAGTGTACAGCTT 1158
Db 1273 TACAAGATGATAGCTGTAAGTGTTCATGAGATCAACAGTAAACGAGTGTACAGCTT 1332
QY 1159 GTTTTTCGCGCCCTGTTGGCGAGCTACGGATCAATTAATCACTGCTGCAAGTCTCGAAG 1218
Db 1333 GTTTTTCGCGCCCTGTTGGCGAGCTACGGATCAATTAATCACTGCTGCGGTCTCGAAG 1392
QY 1219 AATTTGGACACCCACCTGCTCAGATTTGACCCCTGCTGATCCAGCGCTTGGTGTCA 1278
Db 1393 AATTTGGACACCCACCTGCTCAGATTTGACCCCTGCTGATCCAGCGCTTGGTGTCA 1452
QY 1279 GAGCGTGTGTGCTCCCGCTGTCACACGTTGTGCTTGTGGCTTTTGTGCGACAGT 1338
Db 1453 GAGAGTGTGTGCTCCCGCTGTCACACGTTGTGCTTGTGGCTTTTGTGCGGCAAC 1512
QY 1339 GCCACCAAAACCGAATGGGAGATGCGTACCGCTGCTCAACGCAAGCAGCGCAATGCG 1398
Db 1513 GCCTCCCAAAACCGTATGGGAGGATGCGTACCGCTGCTCAACGCAAGCAGCGCAATGCG 1572
QY 1399 GAGAGGTTCCGAACCGTTTGAAGTTTGGGGGTTGGCGGAGGGCGCTTTGGCGGTG 1458
Db 1573 GAGAGGTTCCGAACCGTTTGAAGTTTGGGGGTTGGCGGAGGGCGCTTTGGCGGTG 1632
QY 1459 AGCCAGCAGGGCAGAAATCAACGCTATCACTTTGCAAAACACGCGTTCAACGCTGGTGCG 1518
Db 1633 AGCCAGCAGGGCAGAAATCAGCGGTATCGTTTTCGAAACACGCGTTCAACGCTGGTGCG 1692
QY 1519 TCGGTACAGTTTCAAGGTTTCGAGCTGTCGAGTCTCTTGTGCTGGGTGCGAGCTGGAC 1578
Db 1693 TCGGTACAGTTTCAAGGTTTCGAGCTGTCGAGTCTCTTGTGCTGGGTGCGAGCTGGAC 1752

QY 1579 TCTTCTGGTGGCAAAAAAATCTCTGGGGCTCTCGTACGACGAGAACCAAGTGGCAGCCA 1638
Db 1753 TCTTCTGGCGCAAAAAAATCTCTGGGGCTCTCGTACGACGAGAACCAAGTGGCAGCCA 1812
QY 1639 ATATACGGATCAACGCGCGGTGACCGGATCGTGGGAGATGGGTAAAGGTACCAAC 1698
Db 1813 ATATACGGATCAACGCGCGGTGACCGGATCGTGGGAGATGGGTAAAGGTACCAAC 1872
QY 1699 GTGGTCTTACGATGGCAATAAAATTTGGTTCGGTGTACATTGATGGAGAACCTCTGGAG 1758
Db 1873 TTGGTCTTACGATGGCAATAAAATTTGGCTCCGTTGATATTGATGGAGAACCTCTGGAG 1932
QY 1759 GGTTCAGGGAGACCCCTTGTGCCAGACGGGAGACCGCTGACATCTCCACCTTACGCTT 1818
Db 1933 GGTTCAGGACAGACCCCTTGTGCCAGACGGGAGACCGCTGACATCTCCACCTTACGCTT 1992
QY 1819 GCGGGTATGGAAGGAGTGTATGCCAACCATTAAGCCAGTGCAGCGTGTGATTAATGTTCTT 1878
Db 1993 GCGGGTATTAAGGAGTGTATGCCAACCATTAAGCCAGTGCAGCGTGTGATTAATGTTCTT 2052
QY 1879 CTTTACAACCGTCACTGATATGCCAGGAGATCAGGACCTTCTTGTGAGCCAGGACCTG 1938
Db 2053 CTTTACAACCGACAGCTGAATACCGAGGATCAGGACCTTCTTGTGAGCCAGGACCTT 2112
QY 1939 ATTGGCAGGAAGCACATG 1959
Db 2113 ATTGGCAGGAAGCACATG 2133

RESULT 6

ADC21532

ID ADC21532 standard; DNA; 5403 BP.

AC ADC21532;
XXDT 18-DEC-2003 (first entry)
XX

DE T. cruzi trans-sialidase gene, TS, clone 7F.

KW Trans-sialidase; gene; TS; ds; neurotrophic peptide; interleukin-6;
IL-6 secretion inducing peptide; neuron; glial cell; trophic support;
ciliary neurotrophic factor; CNTF; leukemia inhibitory factor; LIF;
amyotrophic lateral sclerosis; Alzheimer's disease; Parkinson's disease;
Huntington's disease; Chagas' disease; peripheral neuropathy; palsy;
multiple sclerosis; stroke; brain trauma; spinal cord trauma;
peripheral nerve trauma.

OS Trypanosoma cruzi.

Key Location/Qualifiers

CDS 484..3972
/*tag= a
/product= "Trans-sialidase"

US2002137667-A1.

PD 26-SEP-2002.

PP 20-DEC-2000; 2000US-00745008.

PR 20-DEC-1999; 99US-0172881P.

PA (TUFT) UNIV TUFTS.

PI Chuenkova M, Pereira MA;

DR WPI; 2003-786654/74.

DR P-PSDB; ADC21533.

PT T. cruzi trans-sialidase derived neurotrophic and interleukin-6 secretion
inducing peptides, useful in the treatment of neuronal degeneration
caused for example by Chagas' disease.

XX Example 1; SEQ ID NO 33; 79pp; English.
XX The invention relates to a T. cruzi trans-sialidase (TS) derived
XX neurotrophic peptide appearing as ADC21513 or ADC21511 called C44 and
XX C14, or its variant, and an interleukin-6 (IL-6) secretion inducing
XX peptide appearing as ADC21531 called TR-1 (terminal repeat 1). Also
XX included are a composition comprising the peptides, fusion protein
XX comprising the peptides and a fusion partner, a composition comprising
XX the fusion protein and a physiological acceptable carrier, providing
XX trophic support for neurons or glial cells in a mammal (comprising
XX administering a therapeutically effective amount of T. cruzi trans-
XX sialidase (TS) or a neurotrophic variant) and stimulating (M3) the
XX secretion of IL-6 (interleukin-6) in a mammal comprising administering TS
XX or an IL-6 inducing variant. The fusion partner comprises a mammalian
XX neurotrophic factor which is ciliary neurotrophic factor (CNTF) or
XX leukemia inhibitory factor (LIF). The peptides are useful in providing
XX trophic support for neurons and glial cells in a mammal suffering a
XX condition selected from: amyotrophic lateral sclerosis, Alzheimer's
XX disease, Parkinson's disease, Huntington's disease, Chagas' disease,
XX peripheral neuropathy, palsies, multiple sclerosis, stroke, brain trauma,
XX spinal cord trauma and peripheral nerve trauma, and in stimulating the
XX secretion of IL-6. The present sequence encodes trans sialidase clone 7F.
XX
SQ Sequence 5403 BP; 1092 A; 1599 C; 1479 G; 1233 T; 0 U; 0 Other;

Query Match 85.0%; Score 1708.2; DB 9; Length 5403;
Best Local Similarity 93.3%; Pred. No. 0;
Matches 1852; Conservative 0; Mismatches 123; Indels 10; Gaps 6;

QY 22 CATCATCAGCAGCGCGCTGGTCCCGCGGAGCCCATATGGCACCCGGATCGAGCCGA 81
DB 308 CTTCCTCTTCGCGATCTGCCCGCAGCGAGCGCGCGTACGCCCTGGCACCGGATCGAGCCGA 367

QY 82 GTTGAGCTGTTAAGCGGCAAGCTCGAAGTGCCTATTGCA--RAGGCGGCAAGTC 138
DB 368 GTTGAG--GGTTAAGCGTAAGAAATTCAGCGTGCCTTTGAAGCAAGGCGCGCAAGTC 426

QY 139 ACCGAGCGGGTGTCCACTGCTCCGCTCCCGCCCTCTGTAATGTGACGGGGTGATG 198
DB 427 ACCGAGCGGGTGTCCACTGCTCCGCTTCGCCCTCTGTAATGTGACGGGGTGATG 486

QY 199 GTTGCCATCCGAGCGCTCGCTACGAAACATCAATGACAACTCCCTCATTTGATACGGTG 258
DB 487 GTTGCCATCCGAGCGCTCGCTACGAAACATCAATGACAACTCCCTCATTTGATACGGTG 546

QY 259 GCGAAGTACAGCTGACGATGGGAGACGTGGGAGACCCCAATGGCCATCAAGACAGT 318
DB 547 GCGAAGTACAGCTGACGATGGGAGACGTGGGAGACCCCAATGGCCATCAAGACAGC 606

QY 319 CGTGCATCGTCTGTTCTCGTGTGGTGGATCCACACAGTGTGTAAGGGCAACAAGCTT 378
DB 607 CGTGTATCGTCTGTTCTCGTGTGGTGGATCCACACAGTGTGTAAGGGCAACAAGCTT 666

QY 379 TACGTCTCGTGTGGAAAGCTACAACAGTTTCGAGAGCTACTGGACGTGCGATGGTATGCG 438
DB 667 TACGTCTCGTGTGGAAAGCTACTAGTTTCGAGAAGCTACTGGTTCGCGATGGTATGCG 726

QY 439 AGAGATGGGATATCTGTTGCGGTGGTGGAGTCAAGAGTCCACTGCGGCGCGCAAG 498
DB 727 AGAGATGGGATATCTGTTGCGGTGGTGGAGTCAAGAGTCCACTGCGGCGCGCAAG 786

QY 499 ATAACCTGCGAGTATCAATGGGAGCGCCCGTGTCACTGAAGGAATTTTTCCCGCGGAA 558
DB 787 ATAACCTGCGAGTATCAATGGGAGCGCCCGTGTCACTGAAGGAATTTTTCCCGCGGAA 846

QY 559 ATGGAAGGAATGACACAAATCAATTTCTTTGGCGGTGAGGTGTGCCATTTGGCGTCC 618
DB 847 ATGGAAGGCATGCACAAATCAATTTCTTTGGCGCGCGGTGTGTGCCATTTGAGCGTCC 906

QY 619 AACGGGAATCTGTGTACCTGTGCGAGTTACGAAACAAAGAGCAAGTTTTTTCGAAG 678
DB 907 AACGGGAATCTGTGTACCTGTGCGAGTTACGAAACAAAGAGCAAGTTTTTTCGAAG 966

QY 679 ATCTTCTACTCGGAACGAGCGGCAAGACGTGAAAGTTTGGGAGGGTAGAGTGATTTT 738
DB 967 ATCTTCTACTCGGAAGATGATGGCAAGACGTGAAAGTTTGGGAAGGGTAGGAGCAATTTT 1026

QY 739 GGTGCTCTCAACCTGTGGCCCTTGATGGGAGGGAGAGCTCATATAACACTCGAGTT 798
DB 1027 GGTGCTCTCAACCTGTGGCCCTTGATGGGAGGGAGAGCTCATATAACACTCGAGTT 1086

QY 799 GACTATCGCCCGCTCTGGTGTACGAGTCCAGTGCATGCGGAATTCGTGGGTGGAGGCT 858
DB 1087 GACTTGGAAAGCCCGCTCTGGTGTACGAGTCCAGTGCATGAGAAACCGTGGGTGGAGGCT 1146

QY 859 GTCGGCAGCGCTCTCACTGTGTGGGCCCCCTCACAAAATCGAAACAGCCCGGAGTCAG 918
DB 1147 GTCGGAAACCGCTCTCGCGTGTGTGGGCCCCCTCACAAAATCGAAACAGCCCGGAGTCAG 1206

QY 919 AGCAGCTTCACTCGCGTGAACCATCGAGGGAATCGGTGTATGCTCTTCACACACCCGCTG 978
DB 1207 AGCAGCTTCACTCGCGTGAACCATCGAGGGAATCGGTGTATGCTCTTCACACACCCGCTG 1266

QY 979 AATTTTAAGGAAGGTGGCTGCGCGACCGACTGAACTCTGGCTGACGGAATAACACGCGC 1038
DB 1267 AATTTTAAGGAAGGTGGCTGCGCGACCGACTGAACTCTGGCTGACGGAATAACACGCGC 1326

QY 1039 ATTTATAACGTTGGGCAAGTATCCATTGGTGTATGAAATTCGCGCTACAGCTCCGCTCTG 1098
DB 1327 ATTTATAACGTTGGGCAAGTATCCATTGGTGTATGAAATTCGCGCTACAGCTCCGCTCTG 1386

QY 1099 TACAAGATGATTAAGCTGTACTGTTTGCATGAGATCAACAGTAACGAGGTGTACAGCTT 1158
DB 1387 TACAAGATGATTAAGCTGTACTGTTTGCATGAGATCAACAGTAACGAGGTGTACAGCTT 1446

QY 1159 GTTTTTGCGCGCTGTGGCGAGTACGAGTCAATTAATCACTGTGCTGAGTCTCTGGAAG 1218
DB 1447 GTTTTTGCGCGCTGTGGCGAGTACGAGTCAATTAATCACTGTGCTGAGTCTCTGGAAG 1506

QY 1219 AATTTGGACAGCCACCTGCTCAGCATTTGACCCCTGCTGATCCAGCCGCTTTCGTCTCA 1278
DB 1507 AATT--GGACAGCCACCTGCTCAGCATTTGACCCCTGCTGATCCAGCCGCTTTCGTCTCA 1565

QY 1279 GAGCGTGTGTGGTCCCGCTGTACACAGGTGGTGTGTTGGCTTTTGTGCGACAGT 1338
DB 1566 GAGAGTGGTGTGGTCCCGCTGTACACAGGTGGTGTGTTGGCTTTTGTGCG--GCAAC 1624

QY 1339 GCCACCAAAACCAATGGGAGATGTACCGCTGCTCAACGCAAGACGCGAAATGCG 1398
DB 1625 GCCTCCAAAACGATGGGAGGAT--CGTACCGCTGCTCAACGCAAGACGCGAAATGCG 1583

QY 1399 GAGAGGGTTCCGAACCGTTTGAAGTTTGGGGGGTTGGCGGAGGGGCGCTTTTGGCCGGTG 1458
DB 1684 GAGAGGGTTCCGAACCGTTTGAAGTTTGGGGGGTTGGCGGAGGGGCGCTTTTGGCCGGTG 1743

QY 1459 AGCAGCAGGGGAGAAATCAACGCTATCACTTTGCAACCAACCGCTTTCAGCTGGTGGCG 1518
DB 1744 AGCAGCAGGGGAGAAATCAACGCTATCACTTTTGCAAAACCAACCGCTTTCAGCTGGTGGCG 1803

QY 1519 TCGGTACCAATTCACAGGTTCCGAGCGTCCGAGTCTTTTGTGGGTGGAGCTGGAC 1578
DB 1804 TCGGTACCAATTCACAGGTTCCGAGGCTCCGAGGCTCCGAGTCTTTTGTGGGTGGAGCTGGAC 1863

QY 1579 TCTTCTGGTGGCAAAAATCTCCTGGGCGCTCTCGTACGACGAGAAAGCAACAGTGGCAGCCA 1638
DB 1864 TCTTCTGGGCGCAAAAATCTCCTGGGCGCTCTCGTACGACGAGAAAGCAACAGTGGCAGCCA 1923

QY 1639 ATATACGATCAACCGCGGTGACCGAGATCGGTGGAGATGGGTAGAGGTATCCAC 1698
DB 1924 ATATACGATCAACCGCGGTGACCGAGATCGGTGGAGATGGGTAGAGGTATCCAC 1983

QY 1699 GTGGTCTTACGATGGCGAATAAATTTGGTTCGGTGTACATTGATGAGAACTCTTCGAG 1758
DB 1984 TTGGTCTTACGATGGCGAATAAATTTGGTTCGGTGTACATTGATGAGAACTCTTCGAG 2043

QY 1759 GGTTCAGGACGACCGTTGTGCAGACGGGAGCGCCTGACATCTCCACTTCTACGTT 1818
 |||||
 Db 2044 GGTTCAGGACGACCGTTGTGCAGACGGGAGCGCCTGACATCTCCACTTCTACGTT 2103
 |||||
 QY 1819 GCGGGTATGGAAGAGTGATATGCCAACCATTAAGCCACGTGACCGTGAATAATGTTCTT 1878
 |||||
 Db 2104 GCGGGTATGGAAGAGTGATATGCCAACCATTAAGCCACGTGACCGTGAATAATGTTCTT 2163
 |||||
 QY 1879 CTTTACAC--CGTCAGCTGAATCCGAGGAGATCAGGACCTTGTCTTGAGCCAGGAC 1935
 |||||
 Db 2164 CTTTACACCGAGCAGCAGCTGAATACCGAGGAGATCAGGACCTTGTCTTGAGCCAGGAC 2223
 |||||
 QY 1936 CTGATTGGCAGCAAGACACACATGCGGACGACGCGCAGCAGTGAAGAGTACGCC 1995
 |||||
 Db 2224 CTTATTGGCAGCAAGACACACATGACAGCAGCAGCAGTGAAGAGTACGCC 2283
 |||||
 QY 1996 GGATC 2000
 |||||
 Db 2284 TCAAC 2288
 |||||

RESULT 7

AAQ49594
 ID AAQ49594 standard; DNA; 500 BP.

XX AC AAQ49594;

XX 25-MAR-2003 (revised)

DT 26-APR-1994 (first entry)

XX DE Encodes protein with trans-sialidase and/or neuramidase activity.

XX XX Trypanosome; trans-sialidase; neuramidase; sleeping sickness;

KW Chagas' Disease; parasite; ss.

XX OS Trypanosoma cruzi.

XX FH Key Location/Qualifiers

FT CDS 1..500

FT /*tag= a

FT /transl_except= pos:445..447, aa:Val

XX W09318787-Al.

XX 30-SEP-1993.

XX 25-MAR-1993; 93WO-US002869.

XX 25-MAR-1992; 92US-00857519.

PR 10-NOV-1992; 92US-00973851.

XX (UYNV) UNIV NEW YORK STATE.

XX Nussenzweig V, Schenkman S, Eichinger D, Vandekerckhove F;

XX WPI; 1993-320452/40.

XX P-PSDB; AAR42014.

XX New trans-sialidase polypeptide(s) isolated from Trypanosoma - used for

FT transferring sialic acid or for treating or preventing trypanostigote

FT infection.

XX Example 10; Fig 17; 130pp; English.

XX This is the nucleotide sequence of the portion of trans-sialidase which

CC imparts trans-sialidase and/or neuramidase activity. (Updated on 25-MAR-

CC 2003 to correct PN field.)

XX Sequence 500 BP; 119 A; 120 C; 147 G; 114 T; 0 U; 0 Other;

SQ

Query Match 24.8%; Score 498.4; DB 2; Length 500;

Best Local Similarity 99.8%; Pred. No. 1e-142;

Matches 499; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 613 GCCTCCAAACGGGAATCTTGTGTACCCCTGTGCAGTTACGAAACAAAAGAACGATTTTTT 672
 |||||
 Db 1 GCCTCCAAACGGGAATCTTGTGTACCCCTGTGCAGTTACGAAACAAAAGAACGATTTTTT 60
 |||||
 QY 673 TCCAAAGATCTTCTACTCGGAAGACGAGGGCAAGACCTGTGAAGTTTCGGGAGGGTAGGAGT 732
 |||||
 Db 61 TCCAAAGATCTTCTACTCGGAAGACGAGGGCAAGACCTGTGAAGTTTCGGGAGGGTAGGAGT 120
 |||||
 QY 733 GATTTTGGCTGTCTCTCAACCTGTGGCCCTTGTAGTGGGAGGGAAGCTCATATAACACT 792
 |||||
 Db 121 GATTTTGGCTGTCTCTCAACCTGTGGCCCTTGTAGTGGGAGGGAAGCTCATATAACACT 180
 |||||
 QY 793 CGAGTTGACTATATCGCCCGCTCTGTGTACGAGTCCAGTGCATATGGGGAATTCGTGGGTG 852
 |||||
 Db 181 CGAGTTGACTATATCGCCCGCTCTGTGTACGAGTCCAGTGCATATGGGGAATTCGTGGGTG 240
 |||||
 QY 853 GAGGCTGTGCGCACGCTCTCAGCTGTGTGGGGCCCTCACAAAATCGAACCGCCCGC 912
 |||||
 Db 241 GAGGCTGTGCGCACGCTCTCAGCTGTGTGGGGCCCTCACAAAATCGAACCGCCCGC 300
 |||||
 QY 913 AGTCAGACGAGCTTCACTGCCCTGACCATTCGAGGGAATCGCTGTATGCTCTTTCACAC 972
 |||||
 Db 301 AGTCAGACGAGCTTCACTGCCCTGACCATTCGAGGGAATCGCTGTATGCTCTTTCACAC 360
 |||||
 QY 973 CCGCTGAATTTTAAGGGAAGTGGCTGCGGACCGACTGAACTCTGGCTGACGGATAAC 1032
 |||||
 Db 361 CCGCTGAATTTTAAGGGAAGTGGCTGCGGACCGACTGAACTCTGGCTGACGGATAAC 420
 |||||
 QY 1033 CAGCCGATTTATAACGTTGGGCAAGTATCCATTGTGTGATGAAAATTCGCCCTACAGCTCC 1092
 |||||
 Db 421 CAGCCGATTTATAACGTTGGGCAAGTATCCATTGTGTGATGAAAATTCGCCCTACAGCTCC 480
 |||||
 QY 1093 GTCTGTACAAGGATGATAA 1112
 |||||
 Db 481 GTCTGTACAAGGATGATAA 500
 |||||

RESULT 8

AAQ49597

ID AAQ49597 standard; DNA; 500 BP.

XX AC AAQ49597;

XX 25-MAR-2003 (revised)

DT 26-APR-1994 (first entry)

XX DE Clone 154 encoding trans-sialidase gene active portion.

XX KW Trypanosome; trans-sialidase; neuramidase; sleeping sickness;

XX Chagas' Disease; parasite; ss.

XX OS Trypanosoma cruzi.

XX FH Key Location/Qualifiers

FT mat_peptide 1..494

FT /*tag= a

FT /note= "has trans-sialidase activity"

XX W09318787-Al.

XX 30-SEP-1993.

XX 25-MAR-1993; 93WO-US002869.

XX 25-MAR-1992; 92US-00857519.

PR 10-NOV-1992; 92US-00973851.

XX (UYNV) UNIV NEW YORK STATE.

XX Nussenzweig V, Schenkman S, Eichinger D, Vandekerckhove F;

XX WPI; 1993-320452/40.

DR P-PSDB; AAR42017.
XX New trans-sialidase polypeptide(s) isolated from Trypanosoma - used for
PT transferring sialic acid or for treating or preventing trypanostigote
PT infection.
XX
XX
PS Claim 23; Fig 22; 130pp; English.
XX
XX Clones 121 and 151, isolated from T.cruzi DNA library by homology to
CC known neuramidase sequences, were found to have identical sequences in
CC the region of the trans-sialidase gene necessary for enzymatic activity.
CC The 121/151 nucleotide sequence (AAQ49595) is similar to the known TCNA
CC sequence (AAQ49596) but distinct from the sequence of clone 154
XX (AAQ49597). (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 500 BP; 119 A; 120 C; 147 G; 114 T; 0 U; 0 Other;

Query Match 24.8%; Score 498.4; DB 2; Length 500;
Best Local Similarity 99.8%; Pred. No. 1e-142;
Matches 499; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 613 GCGTCCAAACGGGAATCTTGTGTACCTGTGCAGTTACGAACAAAAGAACGAGTTT 672
Db 1 GCGTCCAAACGGGAATCTTGTGTACCTGTGCAGTTACGAACAAAAGAACGAGTTT 60

QY 673 TCCAAAGATCTTCTACTCGGAAGACGAGGGCAAGACGTTGGGAGGAGTAGAGT 732
Db 61 TCCAAAGATCTTCTACTCGGAAGATGATGGCAAGACGTTGGGAGGAGTAGAGT 120

QY 733 GATTTGGCTGCTCTCAACCTGCGGACGAGGCAAGACGTTGGGAGGAGTAGAGT 792
Db 121 GATTTGGCTGCTCTCAACCTGCGGACGAGGCAAGACGTTGGGAGGAGTAGAGT 180

QY 793 CGAGTTGACTATCCGCGCCCTCTGGTGTACGAGTCCAGTGCAGTGGGGAATTCGTGGTG 852
Db 181 CGAGTTGACTATCCGCGCCCTCTGGTGTACGAGTCCAGTGGGGAATTCGTGGTG 240

QY 853 GAGGCTGTGCGGACGCTCTCAACCTGCGGACGAGGCAAGACGTTGGGAGGAGTAGAGT 912
Db 241 GAGGCTGTGCGGACGCTCTCAACCTGCGGACGAGGCAAGACGTTGGGAGGAGTAGAGT 300

QY 913 AGTCAGACGCTCTCAACCTGCGGACGAGGCAAGACGTTGGGAGGAGTAGAGT 972
Db 301 AGTCAGACGCTCTCAACCTGCGGACGAGGCAAGACGTTGGGAGGAGTAGAGT 360

QY 973 CCGCTGAATTTTAAAGGAAGTGGCTGCGGACGAGGCAAGACGTTGGGAGGAGTAGAGT 1032
Db 361 CCGCTGAATTTTAAAGGAAGTGGCTGCGGACGAGGCAAGACGTTGGGAGGAGTAGAGT 420

QY 1033 CAGCGCATTTTAAACGTTGGGCAAGTATCCATTGGTGATGAAATTCGCGCTACAGTCC 1092
Db 421 CAGCGCATTTTAAACGTTGGGCAAGTATCCATTGGTGATGAAATTCGCGCTACAGTCC 480

QY 1093 GTCCTGTACAGGATGATAA 1112
Db 481 GTCCTGTACAGGATGATAA 500

RESULT 9
AAQ49595
ID AAQ49595 standard; DNA; 500 BP.
XX
XX AAQ49595;
XX
XX
DT 25-MAR-2003 (revised)
DT 26-APR-1994 (first entry)
XX
XX Clone 121/151 encoding trans-sialidase gene active portion.
XX
XX Trypanosome; trans-sialidase; neuramidase; sleeping sickness;
KW Chagas' Disease; parasite; ss.
XX
OS Trypanosoma cruzi.

XX Key Location/Qualifiers
FH mat_peptide 1..494
FT /*tag= a
FT /note= "has trans-sialidase activity"
XX
XX WO9318787-A1.
XX
XX 30-SEP-1993.
XX
XX 25-MAR-1993; 93WO-US002869.
XX
XX 25-MAR-1992; 92US-00857519.
XX 10-NOV-1992; 92US-00973851.
XX
XX (UYN) UNIV NEW YORK STATE.
XX
XX Nussenzweig V, Schenkman S, Eichinger D, Vandekerckhove F;
XX WPI; 1993-320452/40.
XX P-PSDB; AAR42015.
XX
XX New trans-sialidase polypeptide(s) isolated from Trypanosoma - used for
PT transferring sialic acid or for treating or preventing trypanostigote
PT infection.
XX
XX Claim 23; Fig 22; 130pp; English.
XX
XX Clones 121 and 151, isolated from T.cruzi DNA library by homology to
CC known neuramidase sequences, were found to have identical sequences in
CC the region of the trans-sialidase gene necessary for enzymatic activity.
CC The 121/151 nucleotide sequence (AAQ49595) is similar to the known TCNA
CC sequence (AAQ49596) but distinct from the sequence of clone 154
XX (AAQ49597). (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 500 BP; 119 A; 122 C; 147 G; 112 T; 0 U; 0 Other;

Query Match 24.1%; Score 484; DB 2; Length 500;
Best Local Similarity 98.0%; Pred. No. 2.8e-138;
Matches 490; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 613 GCGTCCAAACGGGAATCTTGTGTACCTGTGCAGTTACGAACAAAAGAACGAGTTT 672
Db 1 GCGTCCAAACGGGAATCTTGTGTACCTGTGCAGTTACGAACAAAAGAACGAGTTT 60

QY 673 TCCAAAGATCTTCTACTCGGAAGACGAGGGCAAGACGTTGGGAGGAGTAGAGT 732
Db 61 TCCAAAGATCTTCTACTCGGAAGATGATGGCAAGACGTTGGGAGGAGTAGAGT 120

QY 733 GATTTGGCTGCTCTCAACCTGCGGACGAGGCAAGACGTTGGGAGGAGTAGAGT 792
Db 121 GATTTGGCTGCTCTCAACCTGCGGACGAGGCAAGACGTTGGGAGGAGTAGAGT 180

QY 793 CGAGTTGACTATCCGCGCCCTCTGGTGTACGAGTCCAGTGCAGTGGGGAATTCGTGGTG 852
Db 181 CGAGTTGACTATCCGCGCCCTCTGGTGTACGAGTCCAGTGGGGAATTCGTGGTG 240

QY 853 GAGGCTGTGCGGACGCTCTCAACCTGCGGACGAGGCAAGACGTTGGGAGGAGTAGAGT 912
Db 241 GAGGCTGTGCGGACGCTCTCAACCTGCGGACGAGGCAAGACGTTGGGAGGAGTAGAGT 300

QY 913 AGTCAGACGCTCTCAACCTGCGGACGAGGCAAGACGTTGGGAGGAGTAGAGT 972
Db 301 AGTCAGACGCTCTCAACCTGCGGACGAGGCAAGACGTTGGGAGGAGTAGAGT 360

QY 973 CCGCTGAATTTTAAAGGAAGTGGCTGCGGACGAGGCAAGACGTTGGGAGGAGTAGAGT 1032
Db 361 CCGCTGAATTTTAAAGGAAGTGGCTGCGGACGAGGCAAGACGTTGGGAGGAGTAGAGT 420

QY 1033 CAGCGCATTTTAAACGTTGGGCAAGTATCCATTGGTGATGAAATTCGCGCTACAGTCC 1092
Db 421 CAGCGCATTTTAAACGTTGGGCAAGTATCCATTGGTGATGAAATTCGCGCTACAGTCC 480

```

QY 1093 GTCCTGTACAAGGATGATAA 1112
Db 481 GTCCTGTACAAGGATGATAA 500

RESULT 10
AAQ49596
ID AAQ49596 standard; DNA; 500 BP.
XX AC
XX AAQ49596;
XX 25-MAR-2003 (revised)
DT 26-APR-1994 (first entry)
XX TCNA trans-sialidase gene active portion.
XX Trypanosome; trans-sialidase; neuramidase; sleeping sickness;
KW Chagas' Disease; parasite; ss.
XX Trypanosoma cruzi.
OS
XX FH Key Location/Qualifiers
FT mat_peptide 1..494
FT /*tag= a
FT /note= "has trans-sialidase activity"
XX WO9318787-A1.
XX 30-SEP-1993.
XX 25-MAR-1993; 93WO-US002869.
XX 25-MAR-1992; 92US-00857519.
PR 10-NOV-1992; 92US-00973851.
XX (UUNY ) UNIV NEW YORK STATE.
XX Nussenzweig V, Schenkman S, Eichinger D, Vandekerckhove F;
XX WPI; 1993-320452/40.
DR P-PSDB; AAR42016.
XX New trans-sialidase polypeptide(s) isolated from Trypanosoma - used for
PT transferring sialic acid or for treating or preventing trypanomastigote
PT infection.
XX Claim 22; Fig 22; 130pp; English.
XX Clones 121 and 151, isolated from T. cruzi DNA library by homology to
CC known neuramidase sequences, were found to have identical sequences in
CC the region of the trans-sialidase gene necessary for enzymatic activity.
CC The 121/151 nucleotide sequence (AAQ49595) is similar to the known TCNA
CC sequence (AAQ49596) but distinct from the sequence of clone 154
CC (AAQ49597). N.B. the TCNA sequence shown in the SEQ.ID Listing No.5 has
CC the G nucleotide at position 486 missing. (Updated on 25-MAR-2003 to
CC correct PN field.)
XX Sequence 500 BP; 124 A; 120 C; 147 G; 109 T; 0 U; 0 Other;

Query Match 22.8%; Score 458.4; DB 2; Length 500;
Best Local Similarity 94.8%; Pred. No. 2.2e-130;
Matches 474; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 613 GCGTCCAAACGGGAATCTTGTTACCTGTGCGAGGTTACGAACAAAGAGCAAGTTTTT 672
Db 1 GCGTCCAAACGGGAATCTTGTTACCTGTGCGAGGTTACGAACAAAGAGCAAGTTTTT 672
QY 673 TCCAAGATCTTCTACTCGGAAGACGAGCGGCGAAGCTTTGGGAGGAGGAGT 732
Db 61 TCCAAGATCTTCTACTCGGAAGATGATGCAAGAGCTGGAACTTTGGGAGGAGGAGT 120
QY 733 GATTTGGCTGCTCTGAACCTGTGCCCTTGATGGGAGGGAAGCTCATATAAACACT 792
Db 733 GATTTGGCTGCTCTGAACCTGTGCCCTTGATGGGAGGGAAGCTCATATAAACACT 792

121 GATTTGGCTGCTCTGAACCTGTGCCCTTGAGTGGAGGGGAAGCTCATATAAACACC 180
793 CGAGTTGACTATCGCCCGCGCTCTGGTGTACGAGTCCAGTGCACATGGGGAAATCTGGGGTG 852
181 CGAGTTGACTGGAAACGCCGCTCTGGTGTACGAGTCCAGTGCACATGGAGAAACCGTGGGTG 240
853 GAGGCTGTGGGACGCTCTCACGTTGTGGGGCCCTCACCAAAATCGAACACGAGCCGGC 912
241 GAGGCTGTGGAAACCGCTCTGCGGTGTGGGGCCCTCACCAAAATCGAACACGAGCCGGC 300
913 AGTCAGACGAGCTTCACTGCCGTGACATCGAGGGAATGCTGTATGCTCTTCACACAC 972
301 AGTCAGACGAGCTTCACTGCCGTGACATCGAGGGAATGCTGTATGCTCTTCACACAC 360
973 CCCTGAATTTAAGGGAAGGTGGCTGCGCGACCGACTGAACCTCTGGCTGACGGATAAC 1032
361 CCCTGAATTTAAGGGAAGGTGGCTGCGCGACCGACTGAACCTCTGGCTGACGGATAAC 420
1033 CAGCGCATTTATAACGTTGGGCAACTATCCATTTGGTGATGAAATTCGCCCTACAGCTCC 1092
421 CAGCGCATTTATAACGTTGGGCAACTATCCATTTGGTGATGAAATTCGCCCTACAGCTCC 480
1093 GTCCTGTACAAGGATGATAA 1112
Db 481 GTCCTGTACAAGGATGATAA 500

RESULT 11
AAZ29720
ID AAZ29720 standard; DNA; 3033 BP.
XX AC
XX AAZ29720;
XX 27-MAR-2000 (first entry)
XX Modified T. cruzi CRP DNA with C-terminal mammalian DAF gene.
XX T. cruzi complement regulatory protein; CRP; GPI anchor addition site;
KW mammalian decay accelerating factor gene; DAF; plasmid pBC12BI-CRP/DAF;
KW recombinant CRP eukaryotic expression cassette; Chagas' disease; vaccine;
KW ds.
XX Trypanosoma cruzi.
XX FH Key Location/Qualifiers
FT CDS 7..3027
FT /*tag= a
FT /product= "Modified complement regulatory protein"
XX WO9960130-A1.
XX 25-NOV-1999.
XX 18-MAY-1999; 99WO-US010977.
XX 21-MAY-1998; 98US-0086197P.
XX (UYP1-) UNIV PITTSBURGH.
XX Norris KA;
XX WPI; 2000-116315/10.
DR P-PSDB; AAY44455.
XX A new vector encoding Trypanosoma cruzi complement regulatory protein,
XX for treatment of Chaga's disease.
XX Example 4; Page 25-29; 33pp; English.
XX The present sequence encodes modified Trypanosoma cruzi complement
XX regulatory protein. The carboxy terminal end was modified to promote
XX surface production in mammalian cells. The predicted GPI anchor addition
XX site was removed and replaced with the C-terminal sequence of mammalian

```

CC decay accelerating factor gene, which is recognised by mammalian cells.
CC This was used in the construction of the plasmid pBC12BI-CRP/DAP for
CC production of recombinant CRP in mammalian cells. Recombinant CRP
CC produced from host cells can be used as a vaccine to prime the immune
CC system of an animal. Hybridomas secreting monoclonal antibodies
CC recognising CRP are produced. This is used to detect Chagas's disease-
CC related proteins and for treatment of the disease
XX
SQ Sequence 3033 BP; 786 A; 698 C; 897 G; 652 T; 0 U; 0 Other;

Query Match 9.3%; Score 186.6; DB 3; Length 3033;
Best Local Similarity 53.6%; Pred. No. 3.9e-46;
Matches 636; Conservative 0; Mismatches 479; Indels 72; Gaps 9;

QY 577 AATCAATTTCTTGGCGTGCAGGTGTTGCCAATT---GTGGCGTCCAAACGGGAATCTTGTG 633
DB 622 AACAAATTTCTGGTGTGGTGGCGCATGCAATTAAGATGGAGGATGATGTCGTACGTG 681
QY 634 TACCCCTGTGAGGTTACGAACAAAGAAAGCAAGTTTTCCTCAAGATCTTCTACTCGGAA 693
DB 682 CTGCCTATTCAAGCCTTGAAGGATGATGGAAGAGTTGTTTCGTTGGTTATCTCTCGCCAA 741
QY 694 GACGAGGCAAGACGTGGAAGTTTGGGAGGTAGGAGTCAATTTTGGCTGCTCTGAACCT 753
DB 742 AAAAATTTTATGGCTGGGAGTTTTCAAATGGTACGAGTATGAGGATGCATCCAGCCT 801
QY 754 GTGGCCCTTCAGTGGGAGGGGA---AGCTCATCAATAACAACCTGAGTTGACTATCGCCG 810
DB 802 CGCGTTCTTGAATGGAAGAAAGAGCTTATATGATGACGTCGTCGTGATACGGCAGC 861
QY 811 CGTCTGGTTPACGATCCAGTGCATGGGAAATTCGTGGGTGGAGGTCGTGGCAGCGTC 870
DB 862 CGCAGGTTTACAGTCAAGCACCATGGGAAATTTGTGACGCGGAAATACGACACTCTT 921
QY 871 TCACGTGTGTGGGCGCCCTCACCABAATCGAACCCGCGGAGTCAGACGACTCACT 930
DB 922 TCACGGGTGTGGGCGCACTCAGTACACGTGTGGGCGCAGCGCGAGGGTGGCTTCGTC 981
QY 931 CGCGTACCAATCGAGGAATCGCTGTATGCTCTTACACACCCCGCTGAATTTT---AAG 987
DB 982 AGCGCATGATCATGGACAGAAAGTCATCTCTGTCGTCGCGCGGTATATTCTGAGAAG 1041
QY 988 GGAAGTGGCTGCGCGACCGACTGAACCTCTGCTGACGATACCGAGCCATTTTATAAC 1047
DB 1042 GATAAAAAAGAAACGGGCGGACTTCACTTGTGCTGACGACATGACAGCAATTTATGAT 1101
QY 1048 GTTGGCCAGTATCCATTTGTGTGATGAATTCGCGCTACAGCTCGCTGTGTACA----- 1102
DB 1102 GTTGGGCGGATATCCCGCTGTGGGTGAGAACGTGCGCGCCAGCACTCTGCTGTACGCCACG 1161
QY 1103 -----AGGATGATAAGCTGTACTGTGTTG 1125
DB 1162 GTCGAGCCGAGCCATTGAAGAGGAGAACCGAGAGAGAACTGTACTGCCCCG 1221
QY 1126 CATGAGATCAAGTAAACGAGG---TGTCAGCCTTGTGTTTGGCGCCTGGTTGGCGAG 1182
DB 1222 TACGAGTCTGTCAGAGGACCGTAAAGTACCAATTCCTTTTGTGGCTTTGACGAGAAAG 1281
QY 1183 CTACGATCATTAATCAAGTGTGTCAGTCTGGAAGAAATGGGACAGCCACTGTCCAGC 1242
DB 1282 TTGGAGGATATGAGGAAGGTATTGGCTGTCATGGAAGAAAGACGCGCAATTTGGAAG 1341
QY 1243 ATTTGACCCCTCTCTGATCCAGCGCTTCCTGTCGAGAGGTGTTGTGTCGCCGTGTC 1302
DB 1342 GAGTACCGCTGCGGGAATGAAGAATATTTGGCGAGTGCTGTGATGCTGTGAATC 1401
QY 1303 ACCACGGTGGTCTTGTGGCTTTTGTGTCAGAGTGCACCAAAACCGAATGGGAGAT 1362
DB 1402 ACCAAG---GGGCTGGTTGGCTCTTATCAACAAAGTCAACTAAGAACAATGAGTGAC 1458
QY 1363 GCGTACCGTGTGCTCAACGACGACG---GCAAAATGCGAGAGGGTTCCGACGGT--- 1416
DB 1459 GAGTACCTCTGCTGATGCAACCGTTTCACGGGGAAGTGAAGATGCTCTGTATGGTGA 1518

QY 1417 TTGAAGTTTTCGGGGGTGTGGCGAGGGCGCTTTTGGCCGTGAGCCAGCAGGGGCAAAAT 1476
DB 1519 TTGACGTTCAAAGAGACCCGAGAGCAGGGCGAGTGGCTCTTGGCGACATGGGGCAGACT 1578
QY 1477 CAACGGTATCACTTTTGCNAACCAACGCGTTTCACGCTGGTGGCGTGGTGCAGTTCACGAG 1536
DB 1579 GTCCCGTACCACCTTTTGCNAACCAAGTTTACTCTTTGTGGCGACGGTGTCCAATGATAAG 1638
QY 1537 GTTCCG-----AGCGTCGCGAGTCTTTTGTGGTTCGAGCCTTGGACTCTTCTGCT 1587
DB 1639 GCTCCGAGACAGCAGCAGCCCATCCCTTTGATGGGTGTGAGGATGAAGATGCACAA 1698
QY 1588 GGCAGAAACTCTCTGGGCTCTCTGACGAGAGACCAAGCAGTGGCAGCCCAATATACGGA 1647
DB 1699 GGCAGCGTCTTTTGTGCTGTACACCCACGACAAAGAGTGGAGGATAATCTTCAAC 1758
QY 1648 TCACACCGCGGTGACGCGCAGCCGATCGTGGGAGATGGTAAAGGTA 1694
DB 1759 GGCAGTCTGCTGCTGCTGCGCGCATATGAGAAATGTGGAATGGGA 1805

RESULT 12
AAZ29719 standard; DNA; 3253 BP.
XX AC AAZ29719;
XX AC (first entry)
XX 27-MAR-2000 (first entry)
XX T. cruzi complement regulating protein DNA.
XX T. cruzi complement regulating protein; CRP; vaccine;
KW Trypanosoma cruzi strain Y; Chagas' disease;
KW recombinant CRP eukaryotic expression cassette; ds.
XX Trypanosoma cruzi.
XX Location/Qualifiers
FH 235..3246
FT /tag= a
FT /product= "Complement regulatory protein"
FT misc_feature 3163..3173
FT /tag= b
FT /note= "GDS anchor addition sequence"
XX WO960130-A1.
XX 25-NOV-1999.
XX 18-MAY-1999; 99WO-US010977.
XX 21-MAY-1998; 98US-0086197P.
XX (UYPI-) UNIV PITTSBURGH.
XX Norris KA;
XX WPI; 2000-116315/10.
XX P-PSDB; AAY44454.
XX A new vector encoding Trypanosoma cruzi complement regulatory protein,
XX for treatment of Chaga's disease.
XX Claim 1; Page 21-25; 33pp; English.
XX The present sequence encodes Trypanosoma cruzi complement regulatory
XX protein. This is a full length coding sequence from strain Y of T. cruzi.
XX This sequence is used to construct a recombinant T. cruzi CRP eukaryotic
XX expression cassette. Recombinant CRP produced from host cells can be used
XX as a vaccine to prime the immune system of an animal. Hybridomas
XX secreting monoclonal antibodies recognising CRP are produced. This is
XX used to detect Chagas's disease-related proteins and for treatment of the

```
CC disease
XX
SQ Sequence 3253 BP; 845 A; 750 C; 962 G; 696 T; 0 U; 0 Other;

Query Match          9.3%; Score 186.6; DB 3; Length 3253;
Best Local Similarity 53.8%; Pred. No. 4e-46;
Matches 636; Conservative 0; Mismatches 479; Indels 72; Gaps 9;

QY 577 AATCAATTTCTTGGCGGTGAGGTGTTGCCATT---GTGGCTCCAAAGCGGAATCTTTGTG 633
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 850 AAACAAATTTCTGGTGGTGGTGGCGCATTAAGATGAGGATGATGGTGGTACGTG 909
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 634 TACCTGTGCGAGTTACGAACAAAGAAAGAAAGTTTTCACAGATCTTCTACTCGGAA 693
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 910 CTGCTATTCAAGCCTTGAAGGATGATGAAAGGTTGTTTGGTTGTTATCTCGCCAA 969
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 694 GACGAGGGCAAGACCTGGAGTTTGGGAGGGTAGGAGTGAATTTTGGTGTCTCTGAACCT 753
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 970 AAAACTTCTTATGGCTGGGAGTTTCAAAATGGTACGAGTGATGAGGATGATCCAGCCT 1029
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 754 GTGGCCCTTGTAGTGGGAGGGGA---AGCTCATATAAACACTCGAGTTGACTATCGCGC 810
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1030 GCGGTTCTTGAATGAAGGAAAGAGCTTATATGATGACGTGCTGTGATGACGGCAGC 1089
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 811 CTTCTGGTGTACGAGTCCAGTGCATGCGGAATTCGTGGGTGGAGGCTGTGGCAGCCTC 870
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1090 CGCAGGGTTTACAGTCAAGCACCATTGGGGAATTTGTGGACGGAGGAATACGACACTCT 1149
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 871 TCACGTGTGTGGGGCCCTCACCAAAATCGAACCCGCGAGTCTCAGACGAGCTTCACT 930
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1150 TCACGCGTGTGGGCACTCAGTACAGTGTGGGGCAGCGCGCAGGGTGGCTTCGTC 1209
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 931 GCCGTGACCATCGAGGGAATCGGTTATGCTCTTTCACACCCCGCTGAATTTT---AAG 987
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1210 AGCGCGATGATCGATGGACAGAAAGTCACTCTCGTCACTCGGCCGCTATATTCTGAGAG 1269
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 988 GGAAGGTGGCTCGCGACCGACTGAACCTCTGGCTGACGGAATACACGCGCATTTATAAC 1047
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1270 GATAAAGAAAGAACCGGCGCACTTCACTTGTGGCTGACGACATGACGCGAATTTATGAT 1329
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1048 GTTGGGCAAGTATCAATTTGTGTGATGAATAATTCGCGCTTACAGCTCCGCTCTGTACA --- 1102
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1330 GTTGGGCGGATATCCGCTGTGGGTGAGAACTGCGCGCCAGCACTCTGCTGTAGCCAGC 1389
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1103 -----AGGATGATAAGCTGTACTGTTTG 1125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1390 GTCGAAGCGCAGCCATTGAAAGAGGAGAAACCGAAGGAGAGAAACTGTACTGCCCG 1449
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1126 CATGAGATCAACAGTAACGAGG---TGTAAGCCTTGTGTTTGGCGCTGTGTTGGCGAG 1182
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1450 TAGAGGTCTGCTGCAGAGGACGGTAAGTACAACTTGTGTTTGTGGGCTTGACGGAGAG 1509
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1183 CTACGGATCAATTAATCAGTGTGCACTCTGGAAGAATTGGACAGACCCACCTGTCCAGC 1242
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1510 TTGGAGGATATGAGGAAGTATTGGCTGATGAGGAAAGGACCGCAATTCGGAG 1569
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1243 ATTTGACCCCTCTGATCCAGCCGCTTGTGTGACAGCGTGTGTTGGTCCCGCTGTC 1302
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1570 GAGTACCGTTCGCGGAATGAAAGAAATAATTTGGCGAGTGGCTGTGATGCTCGTGAATC 1629
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1303 ACCACGTTGGTCTTGTGGCTTTTGTGCGCAGTGCCACCAACCAAGCAATGGGAGGAT 1362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1630 ACCAAG---GGGCTGTTGGCTTTATCCACACAGTCACTTAAGACACATGGAGTAC 1686
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1363 CGGTACCGTGTGCTCAACCAAGCAGC---GCAATGGCGAGAGGTTTCCGAACGGT--- 1416
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1687 GAGTACCTCTGGGTGAATCAACCGTTTACCGGGGAAGTGGAAAGTGCTCTCTGATGGGA 1746
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1417 TTGAAGTTTGGGGGTTTCGGCGAGGGGCTTTTCGGCGCTGAGCCAGCAGGGCAGAT 1476
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1747 TTGACGTTCAAGAGACCGGACAGGGGGAAGTGGCTGTGTGGCGCATGGGGGAGACT 1806
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1477 CAACGGTATCACTTTTGCAAAACACCGCTTCAAGCTGTGGGTGCGGTGACGATTCACGAG 1536
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


RESULT 14	
AAX81756	
ID AAX81756 standard; DNA; 2446 BP.	
XX AC	
XX AAX81756;	
XX AC	
XX 31-AUG-1999 (first entry)	
XX XX	
DE DNA encoding a T. cruzi antigenic polypeptide.	
XX XX	
KW Trypanosoma cruzi epitope; Trypanosoma cruzi infection; antigen; vaccine; Chagas' disease; ss.	
XX KW	
XX OS Trypanosoma cruzi.	
XX XX	
PN W09931246-A1.	
PD 24-JUN-1999.	
XX PD	
XX PF 04-DEC-1998; 98WO-US025871.	
XX PF	
XX PR 18-DEC-1997; 9TUS-00993674.	
XX PR	
XX XX (CORI-) CORIXA CORP.	
PA Read SG, Skeiky YAW, Lodes MJ, Houghton RL, Smith JM, McNeill PD;	
XX PI	
XX DR WPI; 1999-405035/34.	
XX DR	
XX PT New isolated Trypanosoma cruzi epitopes.	
XX PS Claim 34; Page 60-61; 103pp; English.	
XX FS	
CC The specification describes new Trypanosoma cruzi epitopes. A method for detecting Trypanosoma cruzi infection in a biological sample comprises contacting the sample with a polypeptide comprising an epitope of a TC antigen, or a variant of the antigen that differs only in conservative substitutions and/or modifications and detecting the presence of antibodies that bind to the polypeptide in the sample, thereby detecting TC infection. The TC polypeptides can be used in vaccines for inducing protective immunity against Chagas' disease in a patient. The polypeptides and antibodies can also be used for detecting TC infection.	
CC CC AAX81740-61 represent nucleotides that encode TC epitopes from antigenic polypeptides	
XX CC	
SQ Sequence 2446 BP; 578 A; 559 C; 771 G; 525 T; 0 U; 13 Other;	
Query Match 8.9%; Score 179; DB 2; Length 2446;	
Best Local Similarity 51.3%; Pred.No. 7.6e-44;	
Matches 494; Conservative 0; Mismatches 460; Indels 9; Gaps 3	
QY 587 TTGCGGTGCAGGTTCGCCATTGGCGGTCCAACGGGAATCTTGTTGTACCCTGTGCAGG 646	
Db 529 TCGGTGGCGGAGGATCGGGTGTGTGACGAACGGTTACTTGTGCTTCCCATCAGG 588	
QY 647 TTACCAACAAAGAAGCAAGTTTTTTCACAGATCTTTCTACTCGGAAGACGAGGGCAAGA 706	
Db 589 CAGTAGAAAAGGATCGRAGGAGTGTGTACTGTCCATGCGTTTCAACATGCGTATAGAAG 648	
QY 707 CGTGGAAAGTTTGGGAGGGTAGGAGTGATTTTGGCTGTCTGTGAACCTGTGGCCCTTGAGT 766	
Db 649 CATGCGAGCTCTCGTCCGGTACGACAGGTAGTAACCTGCAAGGAACCATCCATCGCGAAAT 708	
QY 767 GGGAGGGGAGACTCATCATRAACAACCTCGAGTTGACTATCCGCCCGCTGTGTTGACAGT 826	
Db 709 TGGAAGGAAATCTAAATTTTAATACTTCTTCGCGTTCGCGGTACTACGAAGTATTCAGGT 768	
QY 827 CCAGTGACATCGGGAAATTCGTGGGTGGAGGGTGTCCGGCACCGCTCTCACCTGTGTGGGGCC 886	
Db 769 CCTTTGACTCTGGNCAAGTTGSGAAATGAGTTGTAGGCCAATAGTCGCTGTGGGGCA 828	
QY 887 CCTCACCAAATCGAACACAGCCCCGGCAGTCTAGAGCAGCTTCTACTGTCGCTGACCATCAGG 946	

Db 61 ATGGCACCCGGATCGAGCCGAGTTGAGCTGTTTAAGCGCAAGCTCGAGGTGCATTT 120
Qy
Db 121 GAAAGGGGGCGCAAGTCAACGAGCGGGTTGCTCACTCGTTCCGCTCCCGCCCTTGT 180
Db 121 GAAAGGGGGCGCAAGTCAACGAGCGGGTTGCTCACTCGTTCCGCTCCCGCCCTTGT 180
Qy 181 AATGTGGAGGGGTGATGTTGCCATCGCGGAGCTCGCTACGAAACATCCAAATGACAA 240
Db 181 AATGTGGAGGGGTGATGTTGCCATCGCGGAGCTCGCTACGAAACATCCAAATGACAA 240
Qy 241 TCCTCATTTGATACGCTGGCGAAGTACAGCTGGAGAGTGGGAGAGTGGGAGAGCCAA 300
Db 241 TCCTCATTTGATACGCTGGCGAAGTACAGCTGGAGAGTGGGAGAGTGGGAGAGCCAA 300
Qy 301 ATTGCCATCAAGAACAGTCTGATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
Db 301 ATTGCCATCAAGAACAGTCTGATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
Qy 361 GTGAAGGGCAACAAGCTTTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Db 361 GTGAAGGGCAACAAGCTTTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Qy 421 ACCTGCATGCTGATCGAGAGAGTGGGATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
Db 421 ACCTGCATGCTGATCGAGAGAGTGGGATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
Qy 481 TCCACTGCGGGGGCAAGTAACTGCGAGTAACTGCGAGTAACTGCGAGTAACTGCGAGT 540
Db 481 TCCACTGCGGGGGCAAGTAACTGCGAGTAACTGCGAGTAACTGCGAGTAACTGCGAGT 540
Qy 541 GAATTTTTCGCGGGGCAAGTAACTGCGAGTAACTGCGAGTAACTGCGAGTAACTGCGAGT 600
Db 541 GAATTTTTCGCGGGGCAAGTAACTGCGAGTAACTGCGAGTAACTGCGAGTAACTGCGAGT 600
Qy 601 GTTGCCATGCTGGGTCGCAAGGAACTTGTGTAACCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Db 601 GTTGCCATGCTGGGTCGCAAGGAACTTGTGTAACCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Qy 661 AAGCAAGTTTTTCCAGATCTTCTACTCGGAAGACGAGGCAAGCTGGAAGTTTGGG 720
Db 661 AAGCAAGTTTTTCCAGATCTTCTACTCGGAAGACGAGGCAAGCTGGAAGTTTGGG 720
Qy 721 GAGGTAGGAGTATTTGGCTGCTGCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Db 721 GAGGTAGGAGTATTTGGCTGCTGCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Qy 781 ATCAATAACACTCGAGTTGATCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
Db 781 ATCAATAACACTCGAGTTGATCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
Qy 841 AATTCGTGGGTGGAGGCTGCGCAAGCTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
Db 841 AATTCGTGGGTGGAGGCTGCGCAAGCTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
Qy 901 AACCGCCCGGCTGAGAGGCTTCACTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Db 901 AACCGCCCGGCTGAGAGGCTTCACTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Qy 961 CTCTTCACACACCGCTGAAATTTAAGGGAAGTGGCTGCGGACCGAGCTGAACTCTGG 1020
Db 961 CTCTTCACACACCGCTGAAATTTAAGGGAAGTGGCTGCGGACCGAGCTGAACTCTGG 1020
Qy 1021 CTGACGGATACAGCGCATTTTAAAGTGGGCAAGTATCCATGCTGCTGCTGCTGCTGCTGCT 1080
Db 1021 CTGACGGATACAGCGCATTTTAAAGTGGGCAAGTATCCATGCTGCTGCTGCTGCTGCTGCT 1080
Qy 1081 GCCTACAGCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
Db 1081 GCCTACAGCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
Qy 1141 AACGAGGTGACAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200

Db 1141 AACGAGGTGACAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
Qy 1201 GTGCTGAGTCTCTGGAAGAAATTTGGGACAGCCACTGTCCAGACATTTGGACCCCTGCTGAT 1260
Db 1201 GTGCTGAGTCTCTGGAAGAAATTTGGGACAGCCACTGTCCAGACATTTGGACCCCTGCTGAT 1260
Qy 1261 CCAGCCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
Db 1261 CCAGCCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
Qy 1321 GGGCTTTTGTGCGCACAGTGCACCAAAACCGAATGGAGAGTGGCTGCTGCTGCTGCTGCTGCTGCT 1380
Db 1321 GGGCTTTTGTGCGCACAGTGCACCAAAACCGAATGGAGAGTGGCTGCTGCTGCTGCTGCTGCTGCT 1380
Qy 1381 GCAAGCACGCGCAAAATCCGAGAGGGTTCGGAACGGTTCGGAAGTTCGCGGGGTTGGCGGA 1440
Db 1381 GCAAGCACGCGCAAAATCCGAGAGGGTTCGGAACGGTTCGGAAGTTCGCGGGGTTGGCGGA 1440
Qy 1441 GGGGCGCTTTGGCGGCTGAGCCAGGAGGAGTTCGGAACGGTTCGGAAGTTCGCGGGGTTGGCGGA 1500
Db 1441 GGGGCGCTTTGGCGGCTGAGCCAGGAGGAGTTCGGAACGGTTCGGAAGTTCGCGGGGTTGGCGGA 1500
Qy 1501 GCGTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
Db 1501 GCGTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
Qy 1561 CTGGGTGCGAGCTGGAATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
Db 1561 CTGGGTGCGAGCTGGAATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
Qy 1621 AAGCACCACTGCGAGCAATATACGATCAACGCGGTTGAGCGGAGTTCGGAACGGTTCGGAAGTTCGCGGGG 1680
Db 1621 AAGCACCACTGCGAGCAATATACGATCAACGCGGTTGAGCGGAGTTCGGAACGGTTCGGAAGTTCGCGGGG 1680
Qy 1681 ATGGTGAAGAGTACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
Db 1681 ATGGTGAAGAGTACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
Qy 1741 GATGAGAACCTCTGAGGGTTCAGGGCAGACCGGTTGTCGAGAGGAGTTCGGAACGGTTCGGAAGTTCGCGGGG 1800
Db 1741 GATGAGAACCTCTGAGGGTTCAGGGCAGACCGGTTGTCGAGAGGAGTTCGGAACGGTTCGGAAGTTCGCGGGG 1800
Qy 1801 ATCTCCCACTTCTAGCTTGGGGGTATGGAAGAGTGTATGTCGAACCATTAAGCCAGCTG 1860
Db 1801 ATCTCCCACTTCTAGCTTGGGGGTATGGAAGAGTGTATGTCGAACCATTAAGCCAGCTG 1860
Qy 1861 ACCTGGAATATGTTCTTCTTAAACCGTCAACCGTCAACCGTCAACCGTCAACCGTCAACCGTCAACCGT 1920
Db 1861 ACCTGGAATATGTTCTTCTTAAACCGTCAACCGTCAACCGTCAACCGTCAACCGTCAACCGTCAACCGT 1920
Qy 1921 TTCTTGAGCCAGGACCTGATTTGGCAGGAGTGTATGTCGAACCATTAAGCCAGCTG 1980
Db 1921 TTCTTGAGCCAGGACCTGATTTGGCAGGAGTGTATGTCGAACCATTAAGCCAGCTG 1980
Qy 1981 GAAAGAGTACGCGCGGATCCGGCTGCTAA 2010
Db 1981 GAAAGAGTACGCGCGGATCCGGCTGCTAA 2010

RESULT 2

US-09-955-909-1

; Sequence 1, Application US/09955909

; Patent No. US20020150995A1

; GENERAL INFORMATION:

; APPLICANT: PELLETIER, Marc

; BARKER, William A.

; ZOPF, David J.

; TITLE OF INVENTION: METHODS FOR PRODUCING

; STALYLOLIGOSACCHARIDES IN A DAIRY SOURCE

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: PENNIE & EDMONDS LLP

/ STREET: 1155 Avenue of the Americas
/ CITY: New York
/ STATE: NY
/ COUNTRY: USA
/ ZIP: 10036-2711
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/955,909
/ FILING DATE: 18-Sep-2001
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/911,393
/ FILING DATE: <Unknown>
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Coruzzi, Laura A
/ REGISTRATION NUMBER: 30,742
/ REFERENCE/DOCKET NUMBER: 7188-032-999
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212)7909090
/ TELEFAX: (212)8699741
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-955-909-1

Query Match 95.8%; Score 1926.2; DB 9; Length 3183;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1931; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 62 TGGCACCCGGATCGAGCCGAGTTGAGCTGTTTAAAGCGGCAAAAGCTCGAAGGTGCCATTG 121
DB 98 TGGCACCCGGATCGAGCCGAGTTGAGCTGTTTAAAGCGGCAAAAGCTCGAAGGTGCCATTG 157
QY 122 AAAAGGGCGCAAGTCAACGAGCGGGTTGTCCACTCGTTCCGCTCCCGCCCTTTGTTA 181
DB 158 AAAAGGGCGCAAGTCAACGAGCGGGTTGTCCACTCGTTCCGCTCCCGCCCTTTGTTA 217
QY 182 ATGTGACCGGGTGATGGTTGCCATCGCGGACGCTCGCTACGAAACATCCAAATGACAACT 241
DB 218 ATGTGACCGGGTGATGGTTGCCATCGCGGACGCTCGCTACGAAACATCCAAATGACAACT 277
QY 242 CCTCATTTGATACGGTGGCGAAGTACAGCGTGGACCATGGGAGAGCTGGGAGACCCAAA 301
DB 278 CCTCATTTGATACGGTGGCGAAGTACAGCGTGGACCATGGGAGAGCTGGGAGACCCAAA 337
QY 302 TTGCCATCAAGAACACTCGTGCATCGTCTGTTTCTCGTGTGGATCCACAGTCAATTG 361
DB 338 TTGCCATCAAGAACACTCGTGCATCGTCTGTTTCTCGTGTGGATCCACAGTCAATTG 397
QY 362 TGAAGGGCAACAAAGCTTTACGTCCTCGTTGGTGAAGTACAAACAGTTCGAGAGCTACTGGA 421
DB 398 TGAAGGGCAACAAAGCTTTACGTCCTCGTTGGTGAAGTACAAACAGTTCGAGAGCTACTGGA 457
QY 422 CGTCCGATCGTGATCGAGAGACTCGGATATTCGCTTGCCTGCTGTGAGGTCAACAACT 481
DB 458 CGTCCGATCGTGATCGAGAGACTCGGATATTCGCTTGCCTGCTGTGAGGTCAACAACT 517
QY 482 CCACTCGGGCGGCAAGATAACTGCGAGTATCAAAATGGGGAGCCCGCTGTCACTGAAG 541
DB 518 CCACTCGGGCGGCAAGATAACTGCGAGTATCAAAATGGGGAGCCCGCTGTCACTGAAG 577
QY 542 AATTTTCCCGCGGAAATGGAGGAATGCACAAATCAATTTCTTGGCGGTGCGAGTG 601

DB 578 AATTTTCCCGCGGAAATGGAGGAATGCACAAATCAATTTCTTGGCGGTGCGAGTG 637
QY 602 TTGCCATTGTGCGTCCAAACGGGAATCTTGTGTACCCCTGTGACAGTTTACGAACAAAAGA 661
DB 638 TTGCCATTGTGCGTCCAAACGGGAATCTTGTGTACCCCTGTGACAGTTTACGAACAAAAGA 697
QY 662 AGCAAGTTTTTCCAAAGATCTTCTACTCGAAGACGAGGGCAAGAGCTGGAAGTTGGGG 721
DB 698 AGCAAGTTTTTCCAAAGATCTTCTACTCGAAGACGAGGGCAAGAGCTGGAAGTTGGGG 757
QY 722 AGGTTAGGAGTGAATTTTGGCTGCTCTGAACCTGTGCCCCTTGAAGTGGAGGGGAAGCTCA 781
DB 758 AGGTTAGGAGTGAATTTTGGCTGCTCTGAACCTGTGCCCCTTGAAGTGGAGGGGAAGCTCA 817
QY 782 TCATAAACACTCGAGTTGACTATCGCCGCGCTCTGGTGTACGAGTCCAGTGAACATGGGA 841
DB 818 TCATAAACACTCGAGTTGACTATCGCCGCGCTCTGGTGTACGAGTCCAGTGAACATGGGA 877
QY 842 ATTCTGGTGGTGGAGGCTGTGCGCACGCTCTCAGCTGTGTGGGGCCCTCAACAAAATCGA 901
DB 878 ATTCTGGTGGTGGAGGCTGTGCGCACGCTCTCAGCTGTGTGGGGCCCTCAACAAAATCGA 937
QY 902 ACCAGCCCGGCACTCAGAGCAGCTTCACTGCCGTGACCATCGAGGGAATCGCGTTTATGC 961
DB 938 ACCAGCCCGGCACTCAGAGCAGCTTCACTGCCGTGACCATCGAGGGAATCGCGTTTATGC 997
QY 962 TCTTTCACACACCCCGCTGAATTTTAAAGGAAGTGGCTGGCGGACCGACTGAACCTCTGGC 1021
DB 998 TCTTTCACACACCCCGCTGAATTTTAAAGGAAGTGGCTGGCGGACCGACTGAACCTCTGGC 1057
QY 1022 TGACGGATAACAGCGCATTTTAAAGTTGGGCAAGTATCCATTGGTGTGATAAATTCG 1081
DB 1058 TGACGGATAACAGCGCATTTTAAAGTTGGGCAAGTATCCATTGGTGTGATAAATTCG 1117
QY 1082 CCTTACAGCTCCGTCCTGTACAAAGGATGAAGCTGTACTGTTTGCATGAGTCAACAGTA 1141
DB 1118 CCTTACAGCTCCGTCCTGTACAAAGGATGAAGCTGTACTGTTTGCATGAGTCAACAGTA 1177
QY 1142 ACGAGGTGTACAGCTTGTGTTTTCGCGCTGTGTTGGCGAGCTACGGATCATTTAAATCAG 1201
DB 1178 ACGAGGTGTACAGCTTGTGTTTTCGCGCTGTGTTGGCGAGCTACGGATCATTTAAATCAG 1237
QY 1202 TGTGTGAGTCTGTGGAAGAAATTGGGACAGCCACTGTGTCCAGCATTTTGCACCCCTGCTGATC 1261
DB 1238 TGTGTGAGTCTGTGGAAGAAATTGGGACAGCCACTGTGTCCAGCATTTTGCACCCCTGCTGATC 1297
QY 1262 CAGCGCTTCTGCTCAGAGCGTGTGTGCTCCGCTGTCCAGCGTTGGTCTTGTGTTG 1321
DB 1298 CAGCGCTTCTGCTCAGAGCGTGTGTGCTCCGCTGTCCAGCGTTGGTCTTGTGTTG 1357
QY 1322 GCTTTTGTGCGCACAGTGCACCAAAACCGAATGGGAGGATGCTACCGCTGCTCAACG 1381
DB 1358 GCTTTTGTGCGCACAGTGCACCAAAACCGAATGGGAGGATGCTACCGCTGCTCAACG 1417
QY 1382 CAAAGCACGCAAAATGCGAGAGGGTTCCGAAACGGTTTGAAGTTTGGCGGGTGGCGGAG 1441
DB 1418 CAAAGCACGCAAAATGCGAGAGGGTTCCGAAACGGTTTGAAGTTTGGCGGGTGGCGGAG 1477
QY 1442 GGGCGTTTGGCGGTGAGCGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1501
DB 1478 GGGCGTTTGGCGGTGAGCGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1537
QY 1502 CGTTTCACTGCTGTGGCGTCCGTTGACGATTTACGAGGTTCCGAGCGTCCGAGTCTTTTGC 1561
DB 1538 CGTTTCACTGCTGTGGCGTCCGTTGACGATTTACGAGGTTCCGAGCGTCCGAGTCTTTGC 1597
QY 1562 TGGGTGCGAGCTGAGCTCTTCTGTGTGGCAAAAACTCTCTGGGCTCTCTGACGAGAGA 1621
DB 1598 TGGGTGCGAGCTGAGCTCTTCTGTGTGGCAAAAACTCTCTGGGCTCTCTGACGAGAGA 1657
QY 1622 AGCACAGTGGCAGCCCAATATACGGATCAACCGCGGTGACCGGACCGGATCGTGGAGA 1681
DB 1658 AGCACAGTGGCAGCCCAATATACGGATCAACCGCGGTGACCGGACCGGATCGTGGAGA 1717

QY 1082 CCTACAGCTCCGCTCTGTACAGGATGATAAGCTGTACTGTTTGCATGAGATCAACAGTA 1141
Db 1025 CCTACAGCTCCGCTCTGTACAGGATGATAAGCTGTACTGTTTGCATGAGATCAACAGTA 1084
QY 1142 ACGAGTGATACAGCTTGTGTTTGGCGCTGTGTGGCGAGCTACGGATCAATTAATCAG 1201
Db 1085 ACGAGTGATACAGCTTGTGTTTGGCGCTGTGTGGCGAGCTACGGATCAATTAATCAG 1144
QY 1202 TGTGTCAGTCTCTGGAAGATTTGGACAGCCACCTGTCCAGCATTTGCACCCCTGCTGATC 1261
Db 1145 TGTGTCAGTCTCTGGAAGATTTGGACAGCCACCTGTCCAGCATTTGCACCCCTGCTGATC 1204
QY 1262 CAGCCGCTTCTGCTCAGAGCGTGTGTGGTCCCGCTGTACACCGTTGGTCTTGTGTTG 1321
Db 1205 CAGCCGCTTCTGCTCAGAGCGTGTGTGGTCCCGCTGTACACCGTTGGTCTTGTGTTG 1264
QY 1322 GCTTTTGTGCGACAGTGCCACCAAAACCGAATGGGAGGATGTCACCGTGTGCTCAACG 1381
Db 1265 GCTTTTGTGCGACAGTGCCACCAAAACCGAATGGGAGGATGTCACCGTGTGCTCAACG 1324
QY 1382 CAGACCGGCAATGCGGAGAGGTTCCGAACGTTTGAAGTTTGGGGGTTGGCGGAG 1441
Db 1325 CAGACCGGCAATGCGGAGAGGTTCCGAACGTTTGAAGTTTGGGGGTTGGCGGAG 1384
QY 1442 GGGCGCTTTGGCGGTTGAGCCAGCAGGGGAGGAATCAACGGTATCACTTTTGCAAAACCAACG 1501
Db 1385 GGGCGCTTTGGCGGTTGAGCCAGCAGGGGAGGAATCAACGGTATCACTTTTGCAAAACCAACG 1444
QY 1502 CGTTACCGTGTGGGCTGTGTCAGATTCAGAGTTCCGAGCTTCGCGAGTCCCTTTCG 1561
Db 1445 CGTTACCGTGTGGGCTGTGTCAGATTCAGAGTTCCGAGCTTCGCGAGTCCCTTTCG 1504
QY 1562 TGGGTGCGAGCTTGGACTCTTCTGTGGGCAAAAACCTCTGGGGCTCTCGTACGACGAGA 1621
Db 1505 TGGGTGCGAGCTTGGACTCTTCTGTGGGCAAAAACCTCTGGGGCTCTCGTACGACGAGA 1564
QY 1622 AGCACAGTGGCAGCCAAATATACGGATCAACGCGGTGACGCCGCGGATCGTGGGAGA 1681
Db 1565 GGCACAGTGGCAGCCAAATATACGGATCAACGCGGTGACGCCGCGGATCGTGGGAGA 1624
QY 1682 TGGGTAAAGGTACACAGTGTCTTACGATGGCGAATAAAATTTGGTTCGGTGTACATTG 1741
Db 1625 TGGGTAAAGGTACACAGTGTCTTACGATGGCGAATAAAATTTGGTTCGGTGTACATTG 1684
QY 1742 ATGGAGAACCTCTGGAGGGTTACAGGCAGACCGTGTGCCAGAGGAGACCCCTGACA 1801
Db 1685 ATGGAGAACCTCTGGAGGGTTACAGGCAGACCGTGTGCCAGAGGAGACCCCTGACA 1744
QY 1802 TCTCCCACTCTTACGTTGGCGGTATGGAAGGAGTGTATGCCAACCAATAGCCAGCTGA 1861
Db 1745 TCTCCCACTCTTACGTTGGCGGTATGGAAGGAGTGTATGCCAACCAATAGCCAGCTGA 1804
QY 1862 CGGTGAATATGTTCTTTTAAACCGTCAAGTGTGAATGCCGAGGAGATCAGAGACCTTGT 1921
Db 1805 CGGTGAATATGTTCTTTTAAACCGTCAAGTGTGAATGCCGAGGAGATCAGAGACCTTGT 1864
QY 1922 TCTTGAGCCAGGACCTGTATGGCAGCGAGACACATGGCAGCAGCGGCGAGAGTG 1981
Db 1865 TCTTGAGCCAGGACCTGTATGGCAGCGAGACACATGGCAGCAGCGGCGAGAGTG 1924
QY 1982 AAGA 1986
Db 1925 CCTGA 1929

RESULT 4

US-09-745-008-1
; Sequence 1, Application US/09745008
; Patent No. US20020137667A1
; GENERAL INFORMATION:
; APPLICANT: Chuenkova, Marina
; APPLICANT: Pereira, Miercio A.

; TITLE OF INVENTION: T. Crusi-Derived Neurotrophic Agents and
; FILE OF INVENTION: Methods of Use Therefor
; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09/745,008
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/172,881
; PRIOR FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2133
; TYPE: DNA
; ORGANISM: Trypanosoma cruzi
; US-09-745-008-1

Query Match 87.3%; Score 1753.8; DB 9; Length 2133;
Best Local Similarity 94.3%; Pred. No. 0;
Matches 1831; Conservative 0; Mismatches 107; Indels 3; Gaps 1;
QY 22 CATCATCACAGCAGCGCTGTGTCGCGCGGAGCCATATGGCACCCGATCGAGCCGA 81
Db 193 CTTCTTTTTCGATCTGCCCGCAGCGCGCGCTTGGCACCCGATCGAGCCGA 252
QY 82 GTTGAGCTGTTTAAGCGGCAAGCTCGAAGTGCCTTTGAA---AAGGCGGCAAGTC 138
Db 253 GTTGAGCTGTTTAAGCGGTAAGATTCGACGGTGCCTTTGAAGACAAGGCGGCAAGTC 312
QY 139 ACCGAGCGGTTGTCCACTGCTTCCGCTCCCGCCCTTGTAAATGTGGACGGGTGATG 198
Db 313 ACCGAGCGGTTGTCCACTGCTTCCGCTCCCGCCCTTGTAAATGTGGACGGGTGATG 372
QY 199 GTTGCCATCGCGAGCCTCGCTACGAAACATCAATGACAACTCCCTCATGTATGATCGGTG 258
Db 373 GTTGCCATCGCGAGCCTCGCTACGACACATCAATGACAACTCCCTCATGTATGATCGGTG 432
QY 259 GCGAAGTACAGCGTGGACGATGGGAGAGCTGGGAGACCCAAATTCGCATCAAGAACAGT 318
Db 433 GCGAAGTACAGCGTGGACGATGGGAGAGCTGGGAGACCCAAATTCGCATCAAGAACAGT 492
QY 319 CGTGACCTCTGTTTCTGCTGTGTGGATCCCAAGTGTATTTGAAGGCAACAAGCTT 378
Db 493 CGTGATCTCTGTTTCTGCTGTGTGGATCCCAAGTGTATTTGAAGGCAACAAGCTT 552
QY 379 TAGCTCCTGTTGGAAGCTACACAGTTCGAGAGCTACTGAGAGCTCGCATGTGTATGTCG 438
Db 553 TAGCTCCTGTTGGAAGCTACTATAGTTCGAGAAGCTACTGCTCGTCCGATGTCGTCG 612
QY 439 AGAGACTGGGATATTCGCTTGCCTTGGTGTGAGGTACGAGTCCACTGCGGGGCGCAAG 498
Db 613 AGAGACTGGGATATTCGCTTGCCTTGGTGTGAGGTACGAGTCCACTGCGGGGCGCAAG 672
QY 499 ATAACTGCGAGTATCAAAATGGGGAGCCCGTGTCACTGAAGGAAATTTTCCCGGCGGAA 558
Db 673 ATAACTGCGAGTATCAAAATGGGGAGCCCGTGTCACTGAAGGAAATTTTCCCGGCGAA 732
QY 559 ATGGAAGGAATGACACAAATCAATTTCTTGGCGGTGCAAGTGTGCCATGTGGCGTCC 618
Db 733 ATGGAAGGATGACACAAATCAATTTCTTGGCGGCGCGGTGTGTCATTTAGCGTCC 792
QY 619 AACGGGAATCTGTGTACCTGTGCAAGTTACGAAACAAAAGAGCAAGTTTTTCCAG 678
Db 793 AACGGGAATCTGTGTACCTGTGCAAGTTACGAAACAAAAGAGCAAGTTTTTCTCCAG 852
QY 679 ATCTTCTACTCGGAACAGCAGGCAAGAGCTGGAAGTTTGGGAGGAGTGTAGAGTGTATTT 738
Db 853 ATCTTCTACTCGGAAGATGATGGAAGTTTGGGAGGAGTGTAGAGGAGTATTTT 912
QY 739 GGCTGCTCTGAACCTGTGGCCCTTGTAGTGGGAGGAGCTCATATAAACAACCTCGAGTT 798
Db 913 GGCTGCTCTGAACCTGTGGCCCTTGTAGTGGGAGGAGCTCATATAAACAACCTCGAGTT 972
QY 799 GACTATCGCGCGCTGTGTGTCAGTGTGACATGGGGAATTCGTGGGTGGAGGCT 858


```
Db 973 GACTGGAACGCCGCTCTGGTGTACGAGTCCAGTGCATGGAGAAACCGTGGGTGGAGGCT 1032
QY 859 GTCCGACGCTCTCACGCTGTGTGGGGCCCTCACCAAAATCGAACCCAGCCGCGCAGTCAG 918
Db 1033 GTCCGAAACCGTCTCGGCTGTGGGGCCCTCACCAAAATCGAACCCAGCCGCGCAGTCAG 1092
QY 919 AGCAGCTTCACTGCGCGTACCATCGAGGGAATGCGTGTATGCTCTTTCACACACCCCGCTG 978
Db 1093 AGCAGCTTCACTGCGCGTACCATCGAGGGAATGCGTGTATGCTCTTTCACACACCCCGCTG 1152
QY 979 AATTTTAAGGNAAGTGTGCGGACCGACCTGAACTCTGGCTGACGATGAACACAGGC 1038
Db 1153 AATTTTAAGGNAAGTGTGCGGACCGACCTGAACTCTGGCTGACGATGAACACAGGC 1212
QY 1039 ATTTATAACGTTGGGCAAGTATCAATTGGTGCATGAAATAATCCGCTACAGCTCCGCTCCTG 1098
Db 1213 ATTTATAACGTTGGGCAAGTATCAATTGGTGCATGAAATAATCCGCTACAGCTCCGCTCCTG 1272
QY 1099 TACAAGGATGATAAGCTGTACTGTTTGCATGAGATCAACAGTAAACGAGGTGTACAGCCTT 1158
Db 1273 TACAAGGATGATAAGCTGTACTGTTTGCATGAGATCAACAGTAAACGAGGTGTACAGCCTT 1332
QY 1159 GTTTTTCGCGCCTGTTGGCGAGCTAGGATCAATTAATCAGTCTGCGAGTCTGTGAAG 1218
Db 1333 GTTTTTCGCGCCTGTTGGCGAGCTAGGATCAATTAATCAGTCTGCGAGTCTGTGAAG 1392
QY 1219 AATTGGGACAGCCACCTGTGCAGCATTTTGACCCCTGCTGATCCAGCGCTTCGTGCTCA 1278
Db 1393 AATTGGGACAGCCACCTGTGCAGCATTTTGACCCCTGCTGATCCAGCGCTTCGTGCTCA 1452
QY 1279 GAGCGTGTGTGTGTCGCGTGTCAACACGTTGGTCTTGTGGCTTTTGTGCGCACAGT 1338
Db 1453 GAGAGTGTGTGTGTCGCGTGTCAACACGTTGGTCTTGTGGCTTTTGTGCGGCAAC 1512
QY 1339 GCCACCAAAACCGAATCGGAGGATGCGTACGCTGCTGCTCAACGACGACGCAAAATGCG 1398
Db 1513 GCCTCCCAAAACGATGGGAGGATGCGTACGCTGCTGCTCAACGACGACGCAAAATGCG 1572
QY 1399 GAGAGGTTCCGAACGTTTGAAGTTTGGCGGGGTTGGCGAGGGGCGCTTTGGCGGCTG 1458
Db 1573 GAGAGGTTCCGAACGTTTGAAGTTTGGCGGGGTTGGCGAGGGGCGCTTTGGCGGCTG 1632
QY 1459 AGCAGAGGGGCGAAGATCAACGGTATCACTTTGCAAAACACCGGTTCAAGCTGTGGCG 1518
Db 1633 AGCAGAGGGGCGAAGATCAAGCGTATCGTTTTCGCAAAACACCGTTCACGCTGTGGCG 1692
QY 1519 TCGGTACGATTCACGAGGTTCCGAGCGTCGCGAGTCCCTTTGCGAGGCGGAGCTGCGAC 1578
Db 1693 TCGGTACGATTCACGAGGTTCCGAGCGTCGCGAGTCCCTTTGCGGTCGAGCCTGGAC 1752
QY 1579 TCTTCTGGTGGCAAAACCTCTGGGCTCTCGTACGAGAGAACACCAAGTGGCAGCCA 1638
Db 1753 TCTTCTGGGCGCAAAACCTCTGGGCTCTCGTACGAGAGAACACCAAGTGGCAGCCA 1812
QY 1639 ATATACGATCAACGCGGTGACCGCGACCGGATCGTGGAGATGGGTGAAGGTATCCAC 1698
Db 1813 ATATACGATCAACGCGGTGACCGCGACCGGATCGTGGAGATGGGTGAAGGTATCCAC 1872
QY 1699 GTGGTCTTACGATGGCGAATAAAATTTGGTTTGGTGTACATGATGAGAACTCTCTGGAG 1758
Db 1873 TTGGTCTTACGATGGCGAATAAAATTTGGTGTACATGATGAGAACTCTCTGGAG 1932
QY 1759 GTTTCAGGGCAGACCGTTGTGCGAGACGGGAGGACGCTGACATCTCCCACTTCTACGTT 1818
Db 1933 GGTTCAGGACAGACCGTTGTGCGAGACGGGAGGACGCTGACATCTCCCACTTCTACGTT 1992
QY 1819 GCGGCTGTGAGAGGATGATATGCCAACATTAAGCCACGTGACGCTGAATATGTTCTT 1878
Db 1993 GCGGCTGTATTAAGAGGATGATATGCCAACATTAAGCCACGTGACGCTGAATATGTTCTT 2052
QY 1879 CTTTACAAACCGTGTGATGCGGAGGAGATCAGGACCTTGTCTTGTAGCCAGGACCTG 1938
Db 2053 CTTTACAAACCGAGCTGAATACCGAGGAGATCAGGACCTTGTCTTGTAGCCAGGACCTT 2112
```

```
QY 1939 ATTGGCAGCGAAGCACACATG 1959
Db 2113 ATTGGCAGCGAAGCACACATG 2133
```

RESULT 5

```
US-09-745-008-33
; Sequence 33, Application US/09745008
; Patent No. US20020137667A1
; GENERAL INFORMATION:
; APPLICANT: Chuenkova, Marina
; APPLICANT: Pereira, Miercio A.
; TITLE OF INVENTION: T. Cruzi-Derived Neurotrophic Agents and
; TITLE OF INVENTION: Methods of Use Thereof
; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09/745,008
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/172,881
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; TYPE: DNA
; ORGANISM: Trypanosoma cruzi
US-09-745-008-33
```

```
Query Match 85.0%; Score 1708.2; DB 9; Length 5403;
Best Local Similarity 93.3%; Pred. No. 0;
Matches 1852; Conservative 0; Mismatches 123; Indels 10; Gaps 6;
```

```
QY 22 CATCATCAGCAGCGGCTGTGGCGCGCAGCCATATGGCACCGGATCGAGCCGA 81
Db 308 CTTCTTTCGATCTGCCCCAGCGAGCCGCGTACGCCCTGGCACCGGATCGAGCCGA 367
QY 82 GTTGAGCTGTTTAAACGCGCAAGCTCGAAGGTGCCATTTGAA---AAGGCGCGCAAGTC 138
Db 368 GTTGAG-GGTTTAAACGTAAGATTTCGACGGTGCCTTTGAAGACAAGCGCGCAAGTC 426
QY 139 ACCGAGCGGTTGTCCACTCGTTCCGCTCCCGCCCTTCTGTTAATGTGGAACGCGGTGATG 198
Db 427 ACCGAGCGGTTGTCCACTCGTTCCGCTCCCGCCCTTCTGTTAATGTGGAACGCGGTGATG 486
QY 199 GTTGCCATCGCGGAGCTCGCTACCAAAATCCCAATGACAACTCCCTCATTTGATACGGTG 258
Db 487 GTTGCCATCGCGGAGCTCGCTACCAAAATCCCAATGACAACTCCCTCATTTGATACGGTG 546
QY 259 GCGAAGTACAGCGTGGACGATGGGAGACGTGGGAGACCCAAATTCATCAAGAACAGT 318
Db 547 GCGAAGTACAGCGTGGACGATGGGAGACGTGGGAGACCCAAATTCATCAAGAACAGC 606
QY 319 CGTGATCGTCTGTTTCTCGTGTGTGATCCCAAGTATGTTGAAGGCGCAAGCTT 378
Db 607 CGTGATCGTCTGTTTCTCGTGTGTGATCCCAAGTATGTTGAAGGCGCAAGCTT 666
QY 379 TACGCTCTGTTGGAAGCTACAAACAGTTCGAGGAGCTACTGACGCTCCATGTTGATCGG 438
Db 667 TACGCTCTGTTGGAAGCTACTATAGTTCGAGAGGACTACTGCTGCTGCATGTTGATCGG 726
QY 439 AGAGACTGGGATATTCGCTTGGCTTGGTGAAGTTCAGAAAGTCCATTCGCGGCGGCAAG 498
Db 727 AGAGACTGGGATATTCGCTTGGCTTGGTGAAGTTCAGAAAGTCCATTCGCGGCGGCAAG 786
QY 499 ATAACTCGAGTATCAAAATGGGAGCGCCGCTGCTCACTGAAGGAATTTTCCCGCGGAA 558
Db 787 ATAACTCGAGTATCAAAATGGGAGCGCCGCTGCTCACTGAAGGAATTTTCCCGCGGAA 846
QY 559 ATGGAAGGAATGCACAAATCAATTTCTTGGCGGTGAGGTGTTGCAATTTGTCGCGTCC 618
Db 847 ATGGAAGGATGCACAAATCAATTTCTTGGCGCGCGGCTGTTGCCATTTGTAGCGTCC 906
QY 619 AACGGGAATCTTGTGTACCTGTGAGTTACGAAACAAAAGAACGAAGTTTTTTCCNAG 678
```


Db 907 AACGGGAATCTTGTGTACCCCTGTGCAGGTTTACGAACAAAGGAAGCAAGTTTCTCCAAG 966
QY ATCTCTTACTCGGAAGCAGGAGCAAGAGCTGTGAAGTTTGGGAGAGGTAGGAGTGAATTT 738
Db 967 ATCTCTTACTCGGAAGATGATGCAAGACGTGGAAGTTTGGGAAGGTAGGAGCGAATTT 1026
QY GGCTGCTCTGAACCTGTGCCCCTTGAGTGGAGGGAAGCTCATATAAACACTCGAGTT 798
Db 1027 GGCTGCTCTGAACCTGTGCCCCTTGAGTGGAGGGAAGCTCATATAAACACCCGAGTT 1086
QY GACTATCGCGCGCTCTGTGTACGAGTCCAGTGAACATGGGGAATTCGTGGGTGAGGCT 858
Db 1087 GACTGGAACCGCGCTGTGTGTACGAGTCCAGTGAACATGGAGAAACCGTGGGTGAGGCT 1146
QY GTGCGGACGCTCTCAAGTGTGGGCCCCCTCACCAAAATCGAAACAGCCCGGAGCTAG 918
Db 1147 GTGCGAACGCTCTCGGTGTGTGGGCCCCCTCACCAAAATCGAAACAGCCCGGAGCTAG 1206
QY AGCAGCTTCACCTGCCGTGACCAATCGAGGGAATGCGTGTATGCTCTTCAACACCCGCTG 978
Db 1207 ACAGCTTCACCTGCCGTGACCAATCGAAGGAATGCGTGTATGCTCTTCAACACCCGCTG 1266
QY AATTTTAAGGAAGGTGGCTCGCGACCGACTGAACCTCTGGCTGACGGAATACCAAGCGC 1038
Db 1267 AATTTTAAGGAAGGTGGCTCGCGACCGACTGAACCTCTGGCTGACGGAATACCAAGCGC 1326
QY ATTTATAAGCTTGGGCAAGTATCCATTTGGTGAATAAATTCGCTTACAGCTCCGCTCG 1098
Db 1327 ATTTATAAGCTTGGGCAAGTATCCATTTGGTGAATAAATTCGCTTACAGCTCCGCTCG 1386
QY TACAAGGATGATAAGCTGTACTCTTTGTCATGAGATCAACAGATTAACAGAGGTGTACAGCTT 1158
Db 1387 TACAAGGATGATAAGCTGTACTCTTTGTCATGAGATCAACAGAGGTGTACAGCTT 1446
QY GTTTTTCGCGCTGTGTGGAGCTACGGAATCAATTAATCAGTGTGACAGTCTCTGGAAG 1218
Db 1447 GTTTTTCGCGCTGTGTGGAGCTACGGAATCAATTAATCAGTGTGCGGTCTCTGGAAG 1506
QY AATTGGGACGACCACTGTCCAGCATTTGCACCCCTGCTGATCCAGCGCTTCTGCTGCA 1278
Db 1507 AATT -GGACAGCCACTGTGTCAGCATTTGCACCCCTGCTGATCCAGCGCTTCTGCTGCA 1565
QY GAGCGTGTGTGGTCCCGCTGTCCAGCATTTGCACCGTTGGTCTTTGTTGGCTTTTGTGCG 1338
Db 1566 GAGAGTGTGTGGTCCCGCTGTCCAGCATTTGCACCGTTGGTCTTTGTTGGCTTTTGTGCG -GCAAC 1624
QY GCCACCAAAACCGAATGGAGGATGCTACCGCTGCTGCAACCGAAGCAAGCAATGCG 1398
Db 1625 GCTCCCAAAACGATATGGAGGAT -GCTACCGCTGCTCAACCGCAAGCAAGCAATGCG 1683
QY GAGAGGTTCCGAACGGTTTGAAGTTTGGGGGTTGGCGAGGCGCTTTTGGCCGGTG 1458
Db 1684 GAGAGGTTCCGAACGGTTTGAAGTTTGGGGGTTGGCGAGGCGCTTTTGGCCGGTG 1743
QY AGCCAGCGGGCAGGAATCAACCGGTATCACTTTGCAAAACCGAGTTCACGCTGTGGCG 1518
Db 1744 AGCCAGCGGGCAGGAATCAGCGGTATCGTTTGGCAACCAACCGTTTACGCTGTGGCG 1803
QY TCGGTGCAATTCAGAGGTTCCGAGGTCGAGGTCCTTTTGTGGGTGCGAGCTTGGAC 1578
Db 1804 TCGGTGCAATTCAGAGGTCGAGGTCGAGGTCCTTTTGTGGGTGCGAGCTTGGAC 1863
QY TCTTCTGGTGGCAAAACCTCTGGGCTCTGTCAGGAGAACCAACAGTGGCAGCCA 1638
Db 1864 TCTTCTGGGCGCAAAACCTCTGGGCTCTGTCAGGAGAACCAACAGTGGCAGCCA 1923
QY ATATACGATCAACCGCGGTGACCGCAGCGGATCGTGGAGATGGGTAAAGGTACCAAC 1698
Db 1924 ATATACGATCAACCGCGGTGACCGCAGCGGATCGTGGAGACGGGTAAAGGTACCAAC 1983
QY GTGGTTCTTACGATGCGGAATAAATTTGGTTGCGTGTACATTTGATGAGAACCTCTGGAG 1758

Db 1984 TTGTTCTTACGATGGCGGAATAAATTTGGCTCGTGTACATTTGATGGAGAACTTCTGGAG 2043
QY GTTTCAGGCGAGACCGTTTGTGCGAGCGGAGAGCGCTGACATCTCCCACTTCTACGTT 1818
Db 2044 GGTTCAGGACAGACCGTTTGTGCGAGCGGAGAGCGCTGACATCTCCCACTTCTACGTT 2103
QY GCGGGTATGGAAGGAGTATATGCCAACCATTAAGCAGTGAACGCTGATATGTTCTT 1878
Db 2104 GCGGGTATGGAAGGAGTATATGCCAACCATTAAGCAGTGAACGCTGATATGTTCTT 2163
QY CTTTACAAC - -CGTCAGCTGAATGCCGAGGAGATCAGAGACCTTGTCTTTGAGCCAGGAC 1935
Db 2164 CTTTACAACCGAGCAGCTGAATACCGAGGAGATCAGAGCTTGTCTTTGAGCCAGGAC 2223
QY CTGATTGGCAGCGAAGCACATGTCGAGCAGCGGAGCAGTGAAGAAAGTACGCCC 1995
Db 2224 CTTATTGGCAGCGAAGCACATGTCGAGCAGCGGAGCAGTGAAGAAAGTACGCCC 2283
QY 1996 GGATC 2000
Db 2284 TCAAC 2288
RESULT 6
US-10-726-692-43
; Sequence 43, Application US/10726692
; Publication No. US20040132077A1
; GENERAL INFORMATION:
; APPLICANT: KEIKO, OTSU
; TITLE OF INVENTION: RECOMBINANT POLYPEPTIDES FOR DIAGNOSING INFECTION WITH
; FILE REFERENCE: PNL21311A
; CURRENT APPLICATION NUMBER: US/10/726,692
; CURRENT FILING DATE: 2003-12-04
; PRIOR APPLICATION NUMBER: US 60/430,654
; PRIOR FILING DATE: 2002-04-02
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 43
; LENGTH: 858
; TYPE: DNA
; ORGANISM: T CRUZI
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(858)
US-10-726-692-43
Query Match 24.8%; Score 498.6; DB 17; Length 858;
Best Local Similarity 97.3%; Pred. No. 1.3e-153;
Matches 507; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 1475 ATCAACGGTATCACTTTGCAAAACCGCGCTTCACGCTGTGGGTGCGGTGACGATTCACG 1534
Db 2 ATCAACAGTATCGTTTTCGAAACCGCGCTTCACGCTGTGGGTGCGGTGACGATTCACG 61
QY 1535 AGTTCCGAGCGTCGCGAGTCTCTTCTGGGTGCGAGCTTGAATCTTCTGTGGCAAAA 1594
Db 62 AGTTCCGAGCGTCGCGAGTCTCTTCTGGGTGCGAGCTTGAATCTTCTGTGGCAAAA 121
QY 1595 AACTCTGGGGTCTCGTACGAGAACGACAGTGGCAGCAATATACGATCAACGC 1654
Db 122 AACTCTGGGGTCTCGTACGAGAACGACAGTGGCAGCAATATACGATCAACGC 181
QY 1655 CGGTGACCGCGACCGGATCGTGGGAGATGGTAAAGGTACCACTGTTTACCATGG 1714
Db 182 CGGTGACCGCGACCGGATCGTGGGAGATGGTAAAGGTACCACTGTTTACCATGG 241
QY 1715 CGAATAAATTTGTTTGGTGTATGATGGAGAACCTCTGGAGGGTTTCAAGGCAACCG 1774
Db 242 CGAATAAATTTGTTTGGTGTATGATGGAGAACCTCTGGAGGGTTTCAAGGCAACCG 301
QY 1775 TTGTGCCAGACGGGAGGACCGCTGACATCTCCCATCTTCTACGTTGGCGGTATGGAAGGA 1834

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 8, 2004, 18:38:23 ; Search time 170 Seconds
(without alignments)
6561.478 Million cell updates/sec

Title: US-10-086-913-1
Perfect score: 2010
Sequence: 1 atgggcagcagccatcatca.....cgcccgatccgctgctaa 2010

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/prodata/2/ina/5A COMB.seq.*
2: /cgn2_6/prodata/2/ina/5B COMB.seq.*
3: /cgn2_6/prodata/2/ina/6A COMB.seq.*
4: /cgn2_6/prodata/2/ina/6B COMB.seq.*
5: /cgn2_6/prodata/2/ina/PCTUS COMB.seq.*
6: /cgn2_6/prodata/2/ina/backfileseq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1926.2	95.8	3193	4	US-08-911-393-1
2	1886.6	93.9	1929	4	US-08-911-393-3
3	498.4	24.8	500	5	PCT-US93-02869-3
4	498.4	24.8	500	5	PCT-US93-02869-7
5	494	24.1	500	5	PCT-US93-02869-6
6	446.4	22.2	499	5	PCT-US93-02869-5
7	179	8.9	2446	3	US-08-934-306-17
8	179	8.9	2446	3	US-08-993-674A-17
9	179	8.9	2446	4	US-09-256-976-17
10	66	3.3	1989	4	US-09-207-388-4
11	65	3.2	366	4	US-09-613-303-11
12	65	3.2	366	4	US-10-267-311-11
13	65	3.2	552	4	US-09-230-078A-3
14	65	3.2	1230	4	US-09-613-303-54
15	65	3.2	1230	4	US-10-267-311-54
16	65	3.2	1707	4	US-09-207-388-10
17	65	3.2	2016	3	US-09-198-723A-110
18	65	3.2	2016	3	US-09-198-723A-111
19	65	3.2	2016	4	US-09-684-881-110
20	65	3.2	2016	4	US-09-684-881-111
21	65	3.2	2847	4	US-09-613-303-20
22	65	3.2	2847	4	US-10-267-311-20
23	64.4	3.2	1624	4	US-09-647-224A-13
24	64	3.2	648	3	US-09-198-723A-112
25	64	3.2	648	3	US-09-198-723A-113
26	64	3.2	648	3	US-09-198-723A-114
27	64	3.2	648	3	US-09-198-723A-115

28	64	3.2	648	4	US-09-684-881-112	Sequence 112, App
29	64	3.2	648	4	US-09-684-881-113	Sequence 113, App
30	64	3.2	648	4	US-09-684-881-114	Sequence 114, App
31	64	3.2	648	4	US-09-684-881-115	Sequence 115, App
32	64	3.2	651	3	US-09-198-723A-92	Sequence 92, Appl
33	64	3.2	651	3	US-09-198-723A-93	Sequence 93, Appl
34	64	3.2	651	3	US-09-198-723A-94	Sequence 94, Appl
35	64	3.2	651	3	US-09-198-723A-95	Sequence 95, Appl
36	64	3.2	651	3	US-09-198-723A-96	Sequence 96, Appl
37	64	3.2	651	3	US-09-198-723A-97	Sequence 97, Appl
38	64	3.2	651	3	US-09-198-723A-98	Sequence 98, Appl
39	64	3.2	651	3	US-09-198-723A-99	Sequence 99, Appl
40	64	3.2	651	3	US-09-198-723A-100	Sequence 100, App
41	64	3.2	651	3	US-09-198-723A-101	Sequence 101, App
42	64	3.2	651	4	US-09-684-881-92	Sequence 92, Appl
43	64	3.2	651	4	US-09-684-881-93	Sequence 93, Appl
44	64	3.2	651	4	US-09-684-881-94	Sequence 94, Appl
45	64	3.2	651	4	US-09-684-881-95	Sequence 95, Appl

ALIGNMENTS

RESULT 1
US-08-911-393-1
; Sequence 1, Application US/08911393
; Patent No. 6323008
; GENERAL INFORMATION:
; APPLICANT: PELLETIER, Marc
; APPLICANT: BARKER, William A.
; APPLICANT: HAKES, David J.
; APPLICANT: ZOPF, David A.
; TITLE OF INVENTION: METHODS FOR PRODUCING
; TITLE OF INVENTION: SIALYLLOIGOSACCHARIDES IN A DAIRY SOURCE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA: US/08/911,393
; FILING DATE: 14-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7188-032-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3183 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-911-393-1

Query Match 95.8%; Score 1926.2; DB 4; Length 3183;
Best Local Similarity 99.6%; Pred. No. 0;

Matches 1931; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
QY 62 TGGACCCCGATCGAGCCGAGTTGAGCTGTTTAAAGCGCAAGAGCTCGAAGTGGCCATTG 121
Db      |||
QY 98 TGGACCCCGATCGAGCCGAGTTGAGCTGTTTAAAGCGCAAGAGCTCGAAGTGGCCATTG 157
Db      |||
QY 122 AAAGGGCGGCAAAAGTCAACGAGCGGGTTGTCCACTCGTTCCGCTCCCGCCCTTGTTA 181
Db      |||
QY 158 AAAGGGCGGCAAAAGTCAACGAGCGGGTTGTCCACTCGTTCCGCTCCCGCCCTTGTTA 217
Db      |||
QY 182 ATGTGACCGGGTGATGGTTGTCATCGCGGACGCTCGCTACGAAACATCCAAATGCAACT 241
Db      |||
QY 218 ATGTGACCGGGTGATGGTTGTCATCGCGGACGCTCGCTACGAAACATCCAAATGCAACT 277
Db      |||
QY 242 CCTCATTTGATACGGTGGCGAAGTACAGCGTGGAGATGGGAGACGTGGGAGACCCCAA 301
Db      |||
QY 278 CCTCATTTGATACGGTGGCGAAGTACAGCGTGGAGATGGGAGACGTGGGAGACCCCAA 337
Db      |||
QY 302 TTGCCATCAAGAACAGTCGTGATCGTCTGTTTCTGTTGTTGATCCACAGTGATTG 361
Db      |||
QY 338 TTGCCATCAAGAACAGTCGTGATCGTCTGTTTCTGTTGTTGATCCACAGTGATTG 397
Db      |||
QY 362 TGAAGGGCAAAAGCTTTACGTCCTGTTGTAAGCTACAAAGTTCGAGGAGCTACTGGA 421
Db      |||
QY 398 TGAAGGGCAAAAGCTTTACGTCCTGTTGTAAGCTACAAAGTTCGAGGAGCTACTGGA 457
Db      |||
QY 422 CGTCGATGGTGATCGGAGAGACTGGGATATTCTGCTTGGCGTTGGTGAGTCAAGT 481
Db      |||
QY 458 CGTCGATGGTGATCGGAGAGACTGGGATATTCTGCTTGGCGTTGGTGAGTCAAGT 517
Db      |||
QY 482 CCACGCGGGCGCAAGTAACTCGGAGTATCAAATGGGGAGCCCCGTGTCACTGAAG 541
Db      |||
QY 518 CCACGCGGGCGCAAGTAACTCGGAGTATCAAATGGGGAGCCCCGTGTCACTGAAG 577
Db      |||
QY 542 AATTTTCCCGCGGCAAGTAACTCGGAGTATCAAATGGGGAGCCCCGTGTCACTGAAG 601
Db      |||
QY 578 AATTTTCCCGCGGCAAGTAACTCGGAGTATCAAATGGGGAGCCCCGTGTCACTGAAG 637
Db      |||
QY 602 TTGCCATTTTCCAGATCTTCTACTCGGAAGCGGCAAGATCTTCTGGCGGTGCGAGTG 661
Db      |||
QY 638 TTGCCATTTTGGCTCCAAAGCGGAATCTTGTGTACCTGTGCAAGTTAGCAAAAAGA 697
Db      |||
QY 662 AGCAAGTTTTCACAGATCTTCTACTCGGAAGCGGCAAGATCTTGTGTACCTGTGCAAGTTAGCAAAAAGA 721
Db      |||
QY 698 AGCAAGTTTTCACAGATCTTCTACTCGGAAGCGGCAAGATCTTGTGTACCTGTGCAAGTTAGCAAAAAGA 757
Db      |||
QY 722 AGGTAGAGTGAATTTGGCTGCTCTGAACCTGTGGCCCTTGAGTGGGAGGGGAAGCTCA 781
Db      |||
QY 758 AGGTAGAGTGAATTTGGCTGCTCTGAACCTGTGGCCCTTGAGTGGGAGGGGAAGCTCA 817
Db      |||
QY 782 TCATAAACACTCGAGTTGATATCGCGCGCTGTGTGTACGAGTCCAGTGACATGGGA 841
Db      |||
QY 818 TCATAAACACTCGAGTTGATATCGCGCGCTGTGTGTACGAGTCCAGTGACATGGGA 877
Db      |||
QY 842 ATTCTGGGTGGAGGTGTGCGCACGCTCTCAGCTGTGGGCGCCCTCACCAAAATCGA 901
Db      |||
QY 878 ATTCTGGGTGGAGGTGTGCGCACGCTCTCAGCTGTGGGCGCCCTCACCAAAATCGA 937
Db      |||
QY 902 ACCAGCCGCGAGTCAGAGAGTTCACTGCCGTGACCATCGAGGGAATGCGTGTATGC 961
Db      |||
QY 938 ACCAGCCGCGAGTCAGAGAGTTCACTGCCGTGACCATCGAGGGAATGCGTGTATGC 997
Db      |||
QY 962 TCTTCACACACCCGCTGAATTTTAAAGGAAGTGGCTGGCGACCGAGTGAACCTCTGGC 1021
Db      |||
QY 998 TCTTCACACACCCGCTGAATTTTAAAGGAAGTGGCTGGCGACCGAGTGAACCTCTGGC 1057
Db      |||
QY 1022 TGACGGATAACGAGCGCAATTTAAGCTTTGGGCAAGTATCCATTTGGTGTGAATAATCCG 1081
Db      |||
QY 1058 TGACGGATAACGAGCGCAATTTAAGCTTTGGGCAAGTATCCATTTGGTGTGAATAATCCG 1117
Db      |||
QY 1082 CCTACAGCTCCGCTCTACAGGATGATAGCTGTACTGTTGATGAGATCAACAGTA 1141
Db      |||
QY 1118 CCTACAGCTCCGCTCTCTACAGGATGATAGCTGTACTGTTGATGAGATCAACAGTA 1177
Db      |||
```

```
QY 1142 ACAGAGTGTAACGCTTGTGTTTTTTCGCGCCTGGTTGGCGAGCTACGAGTCAATTAATCAG 1201
Db      |||
QY 1178 ACAGAGTGTAACGCTTGTGTTTTTTCGCGCCTGGTTGGCGAGCTACGAGTCAATTAATCAG 1237
Db      |||
QY 1202 TGCTGCACTCTCGAAGATTTGGGACAGCCACCTGTCCAGCATTTGCAACCCCTGCTGATC 1261
Db      |||
QY 1238 TGCTGCACTCTCGAAGATTTGGGACAGCCACCTGTCCAGCATTTGCAACCCCTGCTGATC 1297
Db      |||
QY 1262 CAGCCGCTTCTGCTCAGAGCGTGTGTTGGTCCCGCTGTCCACAGGTTGGTCTGTTG 1321
Db      |||
QY 1298 CAGCCGCTTCTGCTCAGAGCGTGTGTTGGTCCCGCTGTCCACAGGTTGGTCTGTTG 1357
Db      |||
QY 1322 GCTTTTGTGCACTGAGTCCCAAAAACCAATGGGAGATGCTACCGCTGCAACG 1381
Db      |||
QY 1358 GCTTTTGTGCACTGAGTCCCAAAAACCAATGGGAGATGCTACCGCTGCAACG 1417
Db      |||
QY 1382 CAAACACGGCAATGGGAGAGGTTCCGAAACGTTTGAAGTTTGGCGGGTTGGCGGAG 1441
Db      |||
QY 1418 CAAACACGGCAATGGGAGAGGTTCCGAAACGTTTGAAGTTTGGCGGGTTGGCGGAG 1477
Db      |||
QY 1442 GGGCGCTTTGGCGGTTGAGCCAGCAGGGGCAAGTCAACCGTATCACTTTGCAAAACGAG 1501
Db      |||
QY 1478 GGGCGCTTTGGCGGTTGAGCCAGCAGGGGCAAGTCAACCGTATCACTTTGCAAAACGAG 1537
Db      |||
QY 1502 CGTTCACGCTGTTGGCGTGGTGACGATTCACAGAGTTCCGAGCGTCGCGAGTCTCTTTC 1561
Db      |||
QY 1538 CGTTCACGCTGTTGGCGTGGTGACGATTCACAGAGTTCCGAGCGTCGCGAGTCTCTTTC 1597
Db      |||
QY 1562 TGGGTGCGAGCTGAGCTCTTCTGCTGGCAAAAACCTCTGGGGCTCTCGTACGACGAGA 1621
Db      |||
QY 1598 TGGGTGCGAGCTGAGCTCTTCTGCTGGCAAAAACCTCTGGGGCTCTCGTACGACGAGA 1657
Db      |||
QY 1622 AGCACAGTGGCAGCAATATACGATCAACCGCGGTGACCGCAGCCGATCGTGGGAGA 1681
Db      |||
QY 1658 AGCACAGTGGCAGCAATATACGATCAACCGCGGTGACCGCAGCCGATCGTGGGAGA 1717
Db      |||
QY 1682 TGGGTAAAGAGTTACCACTGTTCTTACGATGGCGAATAAATTTGGTTCGCTGATCATTTG 1741
Db      |||
QY 1718 TGGGTAAAGAGTTACCACTGTTCTTACGATGGCGAATAAATTTGGTTCGCTGATCATTTG 1777
Db      |||
QY 1742 ATGGAGAACCTCTGGAGGTTTCAAGGCGAGACCGTTGTGCCAGAGGGAGGACGCTGACA 1801
Db      |||
QY 1778 ATGGAGAACCTCTGGAGGTTTCAAGGCGAGACCGTTGTGCCAGAGGGAGGACGCTGACA 1837
Db      |||
QY 1802 TCTCCACTTCTAGCTTGGCGGTTATGGAAGGATGATATGCCAAACCAATAGCCACGTA 1861
Db      |||
QY 1838 TCTCCACTTCTAGCTTGGCGGTTATGGAAGGATGATATGCCAAACCAATAGCCACGTA 1897
Db      |||
QY 1862 CGGTGAATTAATGTTCTTCTTACAAACGCTCAGCTGAATGCCGAGGATCAGGACCTTGT 1921
Db      |||
QY 1898 CGGTGAATTAATGTTCTTCTTACAAACGCTCAGCTGAATGCCGAGGATCAGGACCTTGT 1957
Db      |||
QY 1922 TCTTGAAGCAGGACCTGATTTGGGCAAGGACACATATGGGAGCAGCAGCGGACGAGTG 1981
Db      |||
QY 1958 TCTTGAAGCAGGACCTGATTTGGGCAAGGACACACATATGGGAGCAGCAGCGGACGAGTG 2017
Db      |||
QY 1982 AAAGAAGTACGCCGGATC 2000
Db      |||
QY 2018 CCCACAGTACGCCCTCAAC 2036
Db      |||
```

RESULT 2

US-08-911-393-3
; Sequence 3, Application US/08911393
; Patent No. 6323008
; GENERAL INFORMATION:
; APPLICANT: PELLETER, Marc
; APPLICANT: BARKER, William A.
; APPLICANT: HAKES, David J.
; APPLICANT: ZOPF, David A.
; TITLE OF INVENTION: METHODS FOR PRODUCING
; SIALYLOLIGOSACCHARIDES IN A DAIRY SOURCE

NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/911,393
FILING DATE: 14-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7188-032-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)7909090
TELEFAX: (212)8699741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1929 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-911-393-3

Query Match 93.9%; Score 1886.6; DB 4; Length 1929;

Best Local Similarity 98.8%; Pred. No. 0;

Matches 1901; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY	62	TGCGACCGCGATCGAGCGAGTTGAGCTGTTTAAAGCGGCAAGCTCGAAGTGCCATTG	121
DB	5	TGCGACCGCGATCGAGCGAGTTGAGCTGTTTAAAGCGGCAAGCTCGAAGTGCCATTG	64
QY	122	AAAAGGCGGCAAAAGTCACCGAGCGGTTCTGCACTCGCTCCCGCCCTTGTGTA	181
DB	65	AAAAGGCGGCAAAAGTCACCGAGCGGTTCTGCACTCGCTCCCGCCCTTGTGTA	124
QY	182	ATGTGACGCGGGTGATGGTTGCCATCGCGAGCGCTCGCTACGAAACATCAATGCAACT	241
DB	125	ATGTGACGCGGGTGATGGTTGCCATCGCGAGCGCTCGCTACGAAACATCAATGCAACT	184
QY	242	CCCTCATTTGATCGGTGGGGAAGTACAGCGTGACAGATGGGGAGAGCGTGGAGACCCAA	301
DB	185	CCCTCATTTGATCGGTGGGGAAGTACAGCGTGACAGATGGGGAGAGCGTGGAGACCCAA	244
QY	302	TTGCCATCAAGAACAGTCGTGTCCTGTTCTCGTGTGGATCCCAAGTGAATG	361
DB	245	TTGCCATCAAGAACAGTCGTGTCCTGTTCTCGTGTGGATCCCAAGTGAATG	304
QY	362	TGAAGGGCAACAAAGCTTTACGTCCTCGTTGGGAAGCTACAAAGTTGAGGAGCTACTGGA	421
DB	305	TGAAGGGCAACAAAGCTTTACGTCCTCGTTGGGAAGCTACAAAGTTGAGGAGCTACTGGA	364
QY	422	CGTGGATGTTGATGGAGAGCTGGGATATTCGTTGCCGTTGGTGGAGTCAAGAGT	481
DB	365	CGTGGATGTTGATGGAGAGCTGGGATATTCGTTGCCGTTGGTGGAGTCAAGAGT	424
QY	482	CCACTCGGCGGCAAGATAAATCGAGTATCAAAATGGGGGAGCCCGCTGCTCACTGAAG	541
DB	425	CCACTCGGCGGCAAGATAAATCGAGTATCAAAATGGGGGAGCCCGCTGCTCACTGAAG	484

QY	542	AATTTTCCCGCGGAAATGGAAGGATGCACACAAATCAATTTCTTGGCGGTGCAGTG	601
DB	485	AATTTTCCCGCGGAAATGGAAGGATGCACACAAATCAATTTCTTGGCGGTGCAGTG	544
QY	602	TTGCCATTGTGGCGTCCAAACGGGAATCTTGTGTACCTGTGACAGTTTACGAAACAAAGA	661
DB	545	TTGCCATTGTGGCGTCCAAACGGGAATCTTGTGTACCTGTGACAGTTTACGAAACAAAGA	604
QY	662	AGCAAGTTTTTCCAAAGATCTTCTACTCGGAAGACGAGGGCAAGAGCTGGAAGTTTGGG	721
DB	605	AGCAAGTTTTTCCAAAGATCTTCTACTCGGAAGACGAGGGCAAGAGCTGGAAGTTTGGG	664
QY	722	AGGTTAGGAGTGAATTTTGGCTGCTCTGAACCTGTGGCCCTTGTAGTGGAGGGGAAGTCA	781
DB	665	AGGTTAGGAGCGCTTTTGGCTGCTCTGAACCTGTGGCCCTTGTAGTGGAGGGGAAGTCA	724
QY	782	TCATAAACACTCGAGTTGACTATCGCGCGCTGTGGTGTACAGATCCAGTGAACATGGGA	841
DB	725	TCATAAACACTCGAGTTGACTATCGCGCGCTGTGGTGTACAGATCCAGTGAACATGGGA	784
QY	842	ATTCTGTGGTGGAGGCTGTGCGCACGCTCTCACGTGTGTGGGGCCCTCACCAAAATCGA	901
DB	785	ATTCTGTGGTGGAGGCTGTGCGCACGCTCTCACGTGTGTGGGGCCCTCACCAAAATCGA	844
QY	902	ACCAGCCGCGAGTCAAGCAGCTTCACTCCGCTGACCATCGAGGGAATCGGTGTTATGC	961
DB	845	ACCAGCCGCGAGTCAAGCAGCTTCACTCCGCTGACCATCGAGGGAATCGGTGTTATGC	904
QY	962	TCCTTCAACACCCGCTGAATTTTAAAGGAAGTGGCTGCGCGACCGACTCAACTCTGGC	1021
DB	905	TCCTTCAACACCCGCTGAATTTTAAAGGAAGTGGCTGCGCGACCGACTCAACTCTGGC	964
QY	1022	TGACGGATAACACCGCATTTTAAAGTTGGGCAAGTATCCATTTGTTGATGAAATTCG	1081
DB	965	TGACGGATAACACCGCATTTTAAAGTTGGGCAAGTATCCATTTGTTGATGAAATTCG	1024
QY	1082	CCTACAGCTCCGCTCTGTACAGGATGATAGCTGTACTGTTTGCATGAGATCAACAGTA	1141
DB	1025	CCTACAGCTCCGCTCTGTACAGGATGATAGCTGTACTGTTTGCATGAGATCAACAGTA	1084
QY	1142	ACGAGGTGTACAGCTTGTGTTTTCGCGCTGTGTCGAGCTACGATCATTAATCAAGT	1201
DB	1085	ACGAGGTGTACAGCTTGTGTTTTCGCGCTGTGTCGAGCTACGATCATTAATCAAGT	1144
QY	1202	TGCTGACAGTCTCGGAAGAAATGGGACAGCCACCTGTCCAGCATTTGACCCCTGCTGATC	1261
DB	1145	TGCTGACAGTCTCGGAAGAAATGGGACAGCCACCTGTCCAGCATTTGACCCCTGCTGATC	1204
QY	1262	CAGCCGCTTGTGTCAGAGCGTGTGTCCTCCGCTGTCAACAGGTTGGTCTTGTG	1321
DB	1205	CAGCCGCTTGTGTCAGAGCGTGTGTCCTCCGCTGTCAACAGGTTGGTCTTGTG	1264
QY	1322	GCTTTTGTGTCACAGTCCACCAACCGAAATGGGAGGATGCTACCGCTGCTGCAACG	1381
DB	1265	GCTTTTGTGTCACAGTCCACCAACCGAAATGGGAGGATGCTACCGCTGCTGCAACG	1324
QY	1382	CAAGCACGCAATGCGGAGAGGTTCCGAACCGTTTGAAGTTTGGCGGGTGGCGGAG	1441
DB	1325	CAAGCACGCAATGCGGAGAGGTTCCGAACCGTTTGAAGTTTGGCGGGTGGCGGAG	1384
QY	1442	GGGCGCTTTGGCGGTGAGCAGCGGGGAGAAATCAACGTTATCACTTTGCAACACAG	1501
DB	1385	GGGCGCTTTGGCGGTGAGCAGCGGGGAGAAATCAACGTTATCACTTTGCAACACAG	1444
QY	1502	CGTTACAGCTGGTGGCGTGGTACGATTCACGAGGTTCCGAGCTGCGGAGTCTTTTC	1561
DB	1445	CGTTACAGCTGGTGGCGTGGTACGATTCACGAGGTTCCGAGCTGCGGAGTCTTTTC	1504
QY	1562	TGGGTGCGAGCTGGAAGTCTTCTGTGTGCAAAAGTCTCTCTGGGCTCTCGTACGAGGA	1621
DB	1505	TGGGTGCGAGCTGGAAGTCTTCTGTGTGCAAAAGTCTCTCTGGGCTCTCGTACGAGGA	1564
QY	1622	AGCACCAAGTGGAGCGCAATATACGGATCAACGCCGCTGAGCCGAGTCGGGAGTA	1681

Db 1565 GGCACACAGTGGCAGCCAAATATACGGATCAACGCCCGGTGACGCCGACCGGATCGTGGAGA 1624
Qy 1682 TGGGTAAGAGGTACACCGTGGTCTTACGATGGCGAATAAATTTGGTTCCGTTGATACATTG 1741
Db 1625 TGGGTAAGAGGTACACCGTGGTCTTACGATGGCGAATAAATTTGGTTCCGTTGATACATTG 1694
Qy 1742 ATGGAGAACCTCTGAGAGGTTTCAGGGCAGACCGTTGTGCCAGACGGGAGGACGCTGACA 1801
Db 1685 ATGGAGAACCTCTGAGAGGTTTCAGGGCAGACCGTTGTGCCAGACGAGAGGACGCTGACA 1744
Qy 1802 TCTCCACATCTTACGTTGGCGGGTATGGAAGAGTGATATGCCAAACCAATAGCCACGCTGA 1861
Db 1745 TCTCCACATCTTACGTTGGCGGGTATGGAAGAGTGATATGCCAAACCAATAGCCACGCTGA 1804
Qy 1862 CGGTGAATAATGTTCTTCTTTACACCGTCACTGTAATGCCGAGGAGATCAGGACCTTGT 1921
Db 1805 CGGTGAATAATGTTCTTCTTTACACCGTCACTGTAATGCCGAGGAGATCAGGACCTTGT 1864
Qy 1922 TCTTGAGCCAGGACCTGATGGCAGGAAAGCACATGGGCGAGCAGCGGCGGACGAGTG 1981
Db 1865 TCTTGAGCCAGGACCTGATGGCAGGAAAGCACATGGGCGAGGACATGGACGAGCAGGACGAGTG 1924
Qy 1982 AAGA 1986
Db 1925 CCTGA 1929

RESULT 3

PCT-US93-02869-3
; Sequence 3, Application PC/TUS9302869
; GENERAL INFORMATION:
; APPLICANT: NUSSENZWEIG, VICTOR
; APPLICANT: SCHENKMAN, SERGIO
; APPLICANT: VAN DEN KERKOV, PHILIP
; APPLICANT: EICHINGER, Daniel
; TITLE OF INVENTION: TRANS-SIALIDASE AND METHODS OF USE
; TITLE OF INVENTION: AND MAKING THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/02869
; FILING DATE: 19930325
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/857,519
; FILING DATE: 24-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: TOWNSEND, GUY KEVIN
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: NUSSENZWEIG 1A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

PCT-US93-02869-3

Query Match 24.8%; Score 498.4; DB 5; Length 500;
Best Local Similarity 99.8%; Pred. No. 4.6e-145;
Matches 499; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 613 GCCTCCAAACGGGAATCTTGTGTACCTCTGTGCAGGTACGAAACAAAAGAGCAAGTTTTT 672
Db 1 GCCTCCAAACGGGAATCTTGTGTACCTCTGTGCAGGTACGAAACAAAAGAGCAAGTTTTT 60
Qy 673 TCCAAAGATCTTCTACATCGGAAGACGAGGGCAAGACCTGTGAAGTTTCGGGAGGCTAGGAGT 732
Db 61 TCCAAAGATCTTCTACATCGGAAGACGAGGGCAAGACCTGTGAAGTTTCGGGAGGCTAGGAGT 120
Qy 733 GATTTTGGTGTCTCTGAACCTGTGGCCCTTGTAGTGGAGGGGAAGCTCATATAAACACT 792
Db 121 GATTTTGGTGTCTCTGAACCTGTGGCCCTTGTAGTGGAGGGGAAGCTCATATAAACACT 180
Qy 793 CGAGTTGACTATCGCCCGCTCTGTGTACGAGTCCAGTGACATGGGGAATTCGTGGGTG 852
Db 181 CGAGTTGACTATCGCCCGCTCTGTGTACGAGTCCAGTGACATGGGGAATTCGTGGGTG 240
Qy 853 GAGGCTGTCCGACACGCTCTCACGTGTGTGGGCCCCCTCACAAAATCGAAACGAGCCCGGC 912
Db 241 GAGGCTGTCCGACACGCTCTCACGTGTGTGGGCCCCCTCACAAAATCGAAACGAGCCCGGC 300
Qy 913 AGTCAGAGCAGCTTCACTGCCGTGACCATCGAGGGAAATGCGTGTATGCTCTTCAACAC 972
Db 301 AGTCAGAGCAGCTTCACTGCCGTGACCATCGAGGGAAATGCGTGTATGCTCTTCAACAC 360
Qy 973 CCGCTGAATTTAAGGGAAGTGGCTGCGGACCCGACCTCTGCTGCTGACGGATAAC 1032
Db 361 CCGCTGAATTTAAGGGAAGTGGCTGCGGACCCGACCTCTGCTGCTGACGGATAAC 420
Qy 1033 CAGCGCATTTATAACGTTGGGCAAGTATCCATTTGTGTGATGAAAAATTCGCGCTACAGCTCC 1092
Db 421 CAGCGCATTTATAACGTTGGGCAAGTATCCATTTGTGTGATGAAAAATTCGCGCTACAGCTCC 480
Qy 1093 GTCCCTGTACAGGATGATAA 1112
Db 481 GTCCCTGTACAGGATGATAA 500

RESULT 4

PCT-US93-02869-7
; Sequence 7, Application PC/TUS9302869
; GENERAL INFORMATION:
; APPLICANT: NUSSENZWEIG, VICTOR
; APPLICANT: SCHENKMAN, SERGIO
; APPLICANT: VAN DEN KERKOV, PHILIP
; APPLICANT: EICHINGER, Daniel
; TITLE OF INVENTION: TRANS-SIALIDASE AND METHODS OF USE
; TITLE OF INVENTION: AND MAKING THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/02869
; FILING DATE: 19930325
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/857,519
; FILING DATE: 24-MAR-1992

Query Match	22.2%	Score 446.4	DB 5	Length 499
Best Local Similarity	94.6%	Pred. No. 7.3e-129		
Matches 473	Conservative 0	Mismatches 26	Indels 1	Gaps 1
QY	613	GGGTCCACGGGAATCTTTGTGTACCTGTGCAGTTACGAACAAAAGAAGCAAGATTTTT	672	
Db	1	GGGTCCACGGGAATCTTTGTGTACCTGTGCAGTTACGAACAAAAGAAGCAAGATTTTC	60	
QY	673	TCCAAGATCTTCTACTCGGAGACGAGGGCAAGACGTGGAAGTTTGGGAGGGTAGAGT	732	
Db	61	TCCAAGATCTTCTACTCGGAGATGATGGCAAGACGTGGAAGTTTGGGAGGGTAGGAGC	120	
QY	733	GATTTTGGCTCTCTGAAACCTGTGGCCCTTGAGTGGGAGGGAAGCTCATATAAACACT	792	
Db	121	GATTTTGGCTCTCTGAAACCTGTGGCCCTTGAGTGGGAGGGAAGCTCATATAAACACC	180	
QY	793	CGAGTTGACTATCCGCCCGCTCTGGTGTACGAGTCCAGTGACATGGGGAATTCGTGGGTG	852	
Db	181	CGAGTTGACTGGAAACGCCCGCTCTGGTGTACGAGTCCAGTGACATGGAGAACCCGTGGGTG	240	
QY	853	GAGCTGTGGCAGCTCTTCACGTGTGTGGGGCCCTCACCAAAATCGAACACGCCCGGC	912	
Db	241	GAGCTGTGGAAACCGTCTCGCGTGTGTGGGGCCCTCACCAAAATCGAACACGCCCGGC	300	
QY	913	AGTCAGAGCACTTCACTGCCGTGACCATCGAGGGAATGCCGTGTATGCTCTTCACACAC	972	

Db 301 AGTCAGACGAGCTTCACCTGCCGTGACCACTCAGAGGAATGCGTGTGATGCTCTTCACACAC 360
 QY 973 CGCCTGAATTTTAAAGGGAAGTGCCTGCGCAGCAGCTGAACCTCTGGCTGACGGATAAC 1032
 Db 361 CGCCTGAATTTTAAAGGGAAGTGCCTGCGCAGCAGCTGAACCTCTGGCTGACGGATAAC 420
 QY 1033 CAGCGCATTTATAACGTTGGGCAAGTATCCATTGGTGATGAAAAATTCGCGCTACAGCTCC 1092
 Db 421 CAGCGCATTTATAACGTTGGGCAACTATCCATTGGTGATGAAAAATTCGCGCTACAGCTCC 480
 QY 1093 GTCCTGTACAGGATGATAA 1112
 Db 481 GTCCT-TACAAGGATGATAA 499

 RESULT 7
 US-08-834-306-17
 ; Sequence 17, Application US/08834306
 ; Patent No. 6054135
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Lodes, Michael J.
 ; APPLICANT: Houghton, Raymond L.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION
 ; OF INFECTION BY BACTERIAL AGENTS
 ; NUMBER OF SEQUENCES: 65
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/834,306
 ; FILING DATE: 15-APR-1997
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.422C1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 17:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2446 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-834-306-17

 Query Match 8.9%; Score 179; DB 3; Length 2446;
 Best Local Similarity 51.3%; Pred. No. 3.9e-45;
 Matches 494; Conservative 0; Mismatches 460; Indels 9; Gaps 3;

 QY 587 TTGGGGGTGCAGGTGTTGCCATTGGGGTCCACGGGAATCTTGTTGATCCCTGTGACGG 646
 Db 529 TCGGTGGCGGAGGATCGGGTGTGTGACGAAGAACGGTTACTTTGTGTTCCCATGACGG 588
 QY 647 TTACGAACAAAAAGAACGAAGTTTTTTTCCAAAGATCTTCTACTCGGAAGACGAGGCAAGA 706
 Db 589 CAGTAGAAAAGATGGAAGAGAGTGTGTACTGTCTCATCGGTTTCAACATCGGTATAGAAG 648
 QY 707 CQTGGAAGTTTGGGAGGGTAGGAGTGAATTTTGGCTGCTCTGTAACCTGTGGCCCTTTGAGT 766
 Db 649 CATCGGAGCTCTCGTCCGGTAGCAGCAGGTAGTAACTGCAAGGAACCAATCCATCGGAATT 708

Db	1129	TGGAGCGGATNAAGACTGTGTTGAAGAGGTGCGAGGAGTTGGATGAAGCCCTA---AGAA	1185
Qy	1244	TTTGACCCCTGCTGATCCAGCCGCTTCGTGTCAGAGCGTGGTGTGGTCCCGCTGTCA	1303
Db	1186	CGTGCAGATCCACTGCCACTATCACC CGGTGAGAGGGGCATGTGTTGTCGCCATTC	1245
Qy	1304	CCACGGTTGGTCTTGTGGCTTTTGTGTCGCACAGTGCACA CAAACCGAAATGGGAGGATG	1363
Db	1246	TTACTGACGGGCTGTGTGGCTATTGTCTGTGCTGTGCTGCTGGAGTGTGATGTGACG	1305
Qy	1364	CGTACCGCTGCGTCAACGCAAGCACGCAAAATGCGGAGA--GGTTCCTCCAAACGGTTTGA	1420
Db	1306	AGTACCTCTGCGTGAA CGCAACTGTTTCATGGGACGGTGAGAGGGTTCCTCAATGGAGTGA	1365
Qy	1421	AGTTTGGGGGGTTTGGCGAGGGGCGCTTTTGGCGGTGAGCCAGCAGGGCGAGAATCAAC	1480
Db	1366	CGTTTGAAGGACCGGAGACGGGGCGGGGTGGCTGTGTC CCAAGTGGACAGAAATCAAC	1425
Qy	1481	GGTATCACTTTGCAAAACCAACGCTTCACGCTGTTGGGTGCGTGCAGTTCACGAGGTTTC	1540
Db	1426	CGTACCATTTCTTTACACAAACGTTTCACTCTAGTGGTGTATGGCGGTCAATCCAGATAGGC	1485
Qy	1541	CGA	1543
Db	1486	CGA	1488

```

RESULT 9
US-09-256-976-17
; Sequence 17, Application US/09256976
; Patent No. 6419933
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Smith, John M.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION
; OF CRUZI INFECTION
; FILE REFERENCE: 210121.422C3
; CURRENT APPLICATION NUMBER: US/09/256,976
; CURRENT FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 2446
; TYPE: DNA
; ORGANISM: Trypanosoma cruzi
; FEATURE:
; OTHER INFORMATION: Where any n is an unknown nucleic acid
US-09-256-976-17

```

Query Match	8.9%	Score 179	DB 4	Length 2446
Best Local Similarity	51.3%	Pred. No. 3.9e-45		
Matches 494	Conservative 0	Mismatches 460	Indels 9	Gaps 3
QY	587	TTGCGGTTGACGGGTTGCCATCTGCGCTCCAAACGGGAATCTTGTTACCTGTGCAGG	646	
Dbb	529	TCGGTGGCGGNGATCGGTTGTGACGAAGAACGGTTACCTTGTGCTTCCATGCAGG	588	
QY	647	TTACGAACAAAGAACGAAGTTTTTTCCAAAGTCTTCTACTCGGAAGACGAGGGCAAGA	706	
Dbb	589	CAGTAGAAAAAGGATGGAAGGAGTGTGTACTGTCCATGCGTTTCAACATGCGTATAGAAG	648	
QY	707	CGTGGAAAGTTTGGGGAGGTAGAGAGTGATTTTGGCTGCTCTGAACCTGTGGCCCTTGAGT	766	
Dbb	649	CATCGAGACTCTCGTCCGTTACGACGAGTAGTAACTGCAAGAACCATCCATCGCGAATT	708	
QY	767	GGAGGGGAAGCTCATCATAAACATCGAGTTGACTATACGCCCGCTCTGGTGTACGAGT	826	
Dbb	709	TGGAAGGAATCTAAATTTTAATTACTTCTTGGCGCTCGCGCTTACTACGAAGTATTTCAGGT	768	

QY	827	CCAGTGACATGGGGAAATTCGTGGGTGGAGGCTGTGCGCACGCTCTCAGCTGTGTGGGGCC	886
Db	769	CCCTTGACTCTGGGACAAAGTTGGGAAATGAGTGGTAGCCAAATTAGTCGCGTGTGGGGCA	828
QY	887	CTCTACCAAAATCGAACACGCCCGGCGAGTCAGAGCAGCTTCACTCCGTGTGACCATCGAGG	946
Db	829	ACTCGGTATGGTTCGAAAGAGGTTATGCGCTTGGCTTCCACACCGCTAACCATTTGAGG	888
QY	947	GAATCGGTGTTATGCTCTTACACACCCGCTGAATTTTAAAGGAAGGTGGCTGGCGGACC	1006
Db	889	GAAGGGAAGTGTCTGTTTACCGCCAGTGTATTTGAGAGAGAAATGGTAGGGGTC	948
QY	1007	GACTGAACCTCTGGCTGACGGATAACACGCGCATTTATAACGTTTGGGCAAGTATCCATTG	1066
Db	949	GGCTTTCATCTTTGGGTGACGACGGGTGACGCTGTGCATGATGCTGGGCGGATATCCGATG	1008
QY	1067	GTGATGAAATTCGCGCTTACAGCTCCGTCCTGTA---CAAGGATGATAAGCTGTACTGTT	1123
Db	1009	CAGCTGATGACGCTGCTGCCAGTTCCCTGTTGTATAGCAGTGGGGGCAATCTGATTTCCG	1068
QY	1124	TGCATGAGATCAAAGTAAAGAGGTGTACAGCCTTGTTTTTGCGCGCTGTTGGCGGAGC	1183
Db	1069	TGTACGAGATAAGTGTAGGGGTCTACCGTCTGTGCTGTGCACTGATACGCAGC	1128
QY	1184	TACGGATCATTAATCAAGTCTGCAGTCTGTGAAGAAATTGGGACAGCCACCTGTGCCAGCA	1243
Db	1129	TGGAGCGGATAAAGACTGTGTTGAAGAGGTGGCAGGAGTTGGATGAAGCCCTA---AGAA	1185
QY	1244	TTTGCACCCCTGCTGATCCAGCCGCTTCGTGCTCAGAGCGTGGTTGGTCCCGCTGTCA	1303
Db	1186	CGTGCGAGATCCACTGCCACTATCACCCGGGTGAGAGGGGCATGTGTATTCGTCCCATTC	1245
QY	1304	CCACGGTTGGTCTTGTGGCTTTTGTGCGCACAGTGCACCAAAACCGAATGGAGGATG	1363
Db	1246	TTACTGACGGCTTGTGGCTATTGTGCTGTGTCGATGCGGAGTGTGATGTGACG	1305
QY	1364	CGTACCGCTGCTCAACCGCAAGCACGCAAAATCGGAGA---GGGTTCCCAACGGTTTGA	1420
Db	1306	AGTACCTCTGCGTGAAACGCAACTGTTTCATGGGACGGTGAGAGGGTCTCTCAATGGAGTGA	1365
QY	1421	AGTTTGGGGGGTTGGCGGAGGGGCGCTTTGGCGGTGTAGCCAGCAGGGGCGCAAAATCAAC	1480
Db	1366	CGTTTGAAGGACCCGGAGCAGGGCGGGGTGGGCTGTGTGCCAAGTGGACAGAAATCAAC	1425
QY	1481	GGTATCACTTTGCAAACCAACGCGTTACGCTGTGGCGTCCGTTGACGATTTACAGAGGTTTC	1540
Db	1426	CGTACCATTCTTTACACAAACCGTTCACTCTAGTGGTGTATGGCGGTTCATCCACGATAGGC	1485
QY	1541	CGA 1543	
Db	1486	CGA 1488	

```

RESULT 10
US-09-207-388-4
; Sequence 4, Application US/09207388
; Patent No. 6497880
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Jan
; TITLE OF INVENTION: HEAT SHOCK GENES AND PROTEINS FROM
; TITLE OF INVENTION: NEISSERIA MENINGITIDIS, CANDIDA GLABRATA AND ASPERGILLUS
; TITLE OF INVENTION: FUMIGATUS
; FILE REFERENCE: 870109.411
; CURRENT APPLICATION NUMBER: US/09/207,388
; CURRENT FILING DATE: 1998-12-08
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1989
; TYPE: DNA
; ORGANISM: Neisseria meningitidis
US-09-207-388-4

```


